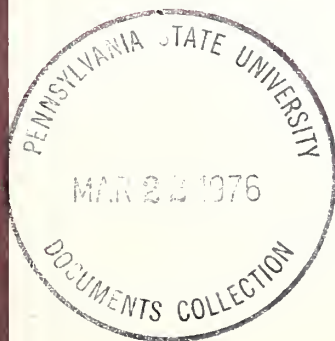


CENSUS TRACT PAPERS
SERIES GE-40, NO. 10

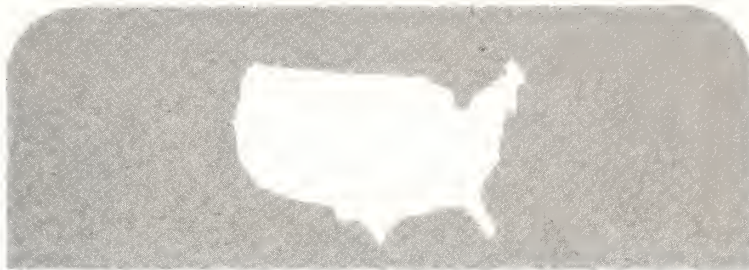
STATISTICAL METHODOLOGY OF REVENUE SHARING AND RELATED ESTIMATE STUDIES

Papers Presented at the Conference
on Small-Area Statistics
American Statistical Association

New York, N.Y.
December 27, 1973



U.S. DEPARTMENT OF COMMERCE
Social and Economic Statistics Administration
BUREAU OF THE CENSUS



CENSUS TRACT PAPERS
SERIES GE-40, NO. 10

**STATISTICAL
METHODOLOGY OF
REVENUE SHARING
AND RELATED
ESTIMATE STUDIES**

**Papers Presented at the Conference
on Small-Area Statistics
American Statistical Association**

New York, N.Y.
December 27, 1973

Issued July 1974



U.S. DEPARTMENT OF COMMERCE
Social and Economic Statistics Administration
BUREAU OF THE CENSUS



BUREAU OF THE CENSUS
Vincent P. Barabba, Director
Robert L. Hagan, Deputy Director
Alva L. Finkner, Associate Director

GEOGRAPHY DIVISION
Morton A. Meyer, Chief

Library of Congress Card No. A 68-7769

SUGGESTED CITATION

U.S. Bureau of the Census, *Census Tract Papers, Series GE-40, No. 10, Statistical Methodology of Revenue Sharing and Related Estimate Studies*, presented at the Conference on Small-Area Statistics, American Statistical Association, New York, N.Y., December 27, 1973. U.S. Government Printing Office, Washington, D.C. 1974.

For sale by the Subscriber Services Section (Publications), U.S. Government Printing Office, Washington, D.C., 20402, and U.S. Department of Commerce district offices. Price \$1.40

PREFACE

The papers published in this report were presented at the Conference on Small-Area Statistics at New York, N.Y. on December 27, 1973 during three sessions of the annual meeting of the American Statistical Association (ASA).

The first two sessions on The Statistical Methodology of Revenue Sharing were sponsored by the ASA Committee on Small-Area Statistics and were organized by Albert Mindlin, Chairman of the Committee. Both were chaired by Sheldon A. Rubin, Liaison for Revenue Sharing, Bureau of the Census, who has provided an introduction. The speakers represented the Bureau of the Census and the Bureau of Economic Analysis of the Social and Economic Statistics Administration, U.S. Department of Commerce and the Office of Revenue Sharing, U.S. Department of Treasury. A discussion paper was provided by Robert P. Strauss, University of North Carolina.

The third session on Recent Developments in Estimation for Local Areas was sponsored jointly by the Social Statistics Section (ASA) and the ASA Committee on Small-Area Statistics. This session was organized and chaired by Paul S. Levy, Associate Professor of Biometry, School of Public Health, University of Illinois Medical Center, Chicago, Illinois who has provided an introduction. The two speakers were from the Bureau of the Census and Temple University. Two discussants represented The Johns Hopkins University and CSR Associates, Alexandria, Virginia.

The report was organized and prepared under the direction of Robert C. Klove, Geographic Research Advisor, Statistical Research Division, Bureau of the Census.

CONTENTS


Page

STATISTICAL METHODOLOGY OF REVENUE SHARING

Introduction	3
Chairman: Sheldon A. Rubin, Bureau of the Census	
Development of Population Estimates for Revenue Sharing Areas	4
Donald E. Starsinic, Bureau of the Census	
Development of the Components of Personal Income and Money Income for the States and Counties	8
Edwin J. Coleman, Bureau of Economic Analysis	
Preparation of Final Revenue Sharing Estimates of Money Income for Political Jurisdictions	18
Roger A. Herriot, Bureau of the Census	
Measurement of State and Local Tax Effort	26
John Coleman, Bureau of the Census	
Statistics of Revenue Sharing After 1 Year	31
Arthur L. Hauser, Office of Revenue Sharing	
Making the New Federalism Work: Goals for General Revenue Sharing's Data Needs .	34
Robert P. Strauss, University of North Carolina	
Formulas for Allocation of Funds Under the General Revenue Sharing Act of 1972 (PL92-512)	37
Prepared by Albert Mindlin, Chairman, Committee on Small-Area Statistics of the American Statistical Association	

RECENT DEVELOPMENTS IN ESTIMATION FOR LOCAL AREAS

Introduction	45
Chairman: Paul S. Levy, University of Illinois	
Use and Evaluation of Synthetic Estimates	46
Maria Elena Gonzalez, Bureau of the Census	
Developments in Statistical Estimation for Local Areas	51
Eugene P. Ericksen, Temple University	
Discussion of Papers by Gonzalez and Ericksen	57
Richard Royall, The Johns Hopkins University	
Discussion of Papers by Gonzalez and Ericksen	59
Hyman B. Kaitz, CSR Associates, Alexandria, Va.	



Digitized by the Internet Archive
in 2012 with funding from
LYRASIS Members and Sloan Foundation

<http://archive.org/details/statisticalmetho00conf>

**Statistical Methodology
of
Revenue Sharing**

Introduction

Sheldon A. Rubin
Bureau of the Census

The ideal Federal program for distribution of funds hopefully starts with the assumption that the basic data used in the distribution formula allow for fair and equitable allocation to all eligible participants. The General Revenue Sharing Program perhaps comes closest of all funding programs to meeting this assumption. Its initial data base consisted of demographic data collected during the 1970 Census of Population and revenue data collected in a special fall 1972 survey of all 38,000 eligible participants. The mission of these two morning sessions of the Conference on Small-Area Statistics of the American Statistical Association is to take a close look at the development of the basic data used by the Office of Revenue Sharing in the allocation of funds, both in the updating process (beyond the 1970 census) and in the case of revenues, those elements that offer the most complexity in developing that key element of the allocation formula.

The panel assembled for the sessions and presenting the papers contained in this publication represent the organizations, staffs and expertise charged legislatively and by agreement with the Office of the Secretary of the Treasury in developing the raw data for the distribution of general revenue sharing funds by the Federal government. The papers provide a clear insight into

the future as far as updating the allocation process is concerned, and a technical look at the revenue developmental process. Students of new federalism and of the Federal funding process will find these papers extremely valuable, both in a critical evaluation of the current program and in contemplation of possibly an improved program.

I am particularly pleased that we were able to add to the distinguished panel the equally distinguished Dr. Robert Strauss of the Department of Economics of the University of North Carolina. Dr. Strauss was instrumental while part of the Office of the Secretary of the Treasury in the development of the legislation and the allocation formula as enacted by Congress for the revenue sharing program.

Enclosed as an appendix are the mathematical derivations of the allocation formulas at the State, county area, and local levels of government. It should be noted that magnetic tapes providing the raw data and resulting allocations are available through the Data Users Service Division of the Bureau of the Census. Additional information about the Census Bureau's participation in the program is also available from that office.

Development of Population Estimates for Revenue Sharing Areas

Donald E. Starsinic, Bureau of the Census

As you all know Federal Revenue Sharing has long been eagerly anticipated by localities throughout the United States for its potential to ease an ever-pressing need for funds. When the program became a reality 2 years ago it was designed to give relief to nearly 39,000 State and local governmental units based on a complex allocation formula. The three principal determinants in calculating the allocation amounts are population, income, and tax effort. In the initial allocations in 1972 both the population and income values used in the allocation formula were obtained from the 1970 Census of Population, while the tax effort was determined by a 1971 survey of government finance conducted by the Bureau of the Census.

As time passes and we get further away from 1970, these sources will become dated rather quickly as bases for the allocations. Obviously some updating procedure is desirable if not essential to avoid the inequities created by changing patterns of population growth, income distribution, and tax effort. With no census likely to provide population and income data for input to the formula until 1980, the administrators of the program are faced with the challenge of creating an adequate updating mechanism based on the best available estimating techniques and data sources. My attempt here will be to describe to you a procedure that we are now testing for updating population levels for all revenue sharing areas, indicating its strengths and pointing out some of the problems in implementing such a procedure, and I will attempt to place it in context with the Census Bureau's ongoing estimates program.

The revenue sharing areas consist of 50 States and the District of Columbia, together with their 3100-odd counties and county equivalents, plus all incorporated places (there are close to 19,000 cities, towns, villages, and boroughs). In addition, in the 21 States where minor civil divisions qualify as revenue sharing governmental units, there are some 17,000 revenue sharing townships (or their equivalent—the town in New England, New York, or Wisconsin). The number of revenue sharing areas changes over time, with the additions resulting from new incorporations and the activation of area governments, and deletions from disincorporations, deactivations, and areas disappearing as the result of annexations or mergers.

For States and counties, population estimates for the 1970's have been prepared either by the Census Bureau or cooperatively with designated State agencies under the Federal-State Cooperative Program for Population Estimates. The Census Bureau's State estimates were used for revenue sharing allocations beginning in 1973, and the Cooperative county estimates are available for use in 1974 allocations.

Below the county level, for the remaining 36,000-odd incorporated places and townships no estimates are now prepared uniformly by any procedure, and to my knowledge no fully effective standard estimating procedure exists to develop population estimates at the small-area level.

The Population Update

To fill this gap and provide population updates for all revenue sharing areas the Census Bureau is now experimentally generating estimates by a component procedure using the Internal Revenue Service's (IRS) master files extracted from the individual income tax 1040 and 1040A returns to generate the migration component. Estimates will be prepared separately for parts of incorporated places crossing county or revenue-sharing township lines.

The basic estimating approach is quite simple and straightforward. To the 1970 population base for an area will be added (1) an estimate of natural increase based, as much as possible, on reported birth and death statistics plus (2) an estimate of net migration based on the IRS files. Some adjustment will be made when necessary for changes in special populations not reflected in the IRS file.

The 1970 base file reflects all population corrections made to the 1970 census. It also incorporates changes due to new incorporations, disincorporations, and large annexations since the 1970 census. This geographic updating is accomplished largely through an annual boundary and annexation survey conducted by the Bureau. Results of the survey as of December 31, 1972 will be reflected in the mid-year 1974 allocations. Adjustments will be made from annexations of at least 250 people and 5 percent of the 1970 place population. Annexations that do not meet this required level are accumulated until the level is reached, then credited to the place. Also certain "unusual annexations" may be considered by the Office of Revenue Sharing at any time regardless of the 250 and 5 percent rule.

For the natural increase component annual births and deaths will be compiled from State vital statistics offices for as many areas as are available. This presently includes over 8,000 cities and about 7,000 townships. These will be supplemented by data from the National Center for Health Statistics for about 300 cities of 10,000 or more not covered by the State agencies. County data are obtainable from all States, and most States tabulate data for the larger cities. A few States have data for all cities and townships.

In those States where these data are not available for all areas to be estimated within a given county, the reported births and deaths for the balance of county will be allocated to those revenue sharing areas in the county not specifically reported on the basis of their share of the 1970 census population. Thus, we assume that the crude birth and death rates are the same in each place in the balance of the county. For later years in the decade we would propose to distribute reported births and deaths in the county balance by the latest available estimated populations under the program for these unspecified places. Vital statistics for institutions and military installations will be reviewed for consistency in handling from area to area as well as consistency over time.

Net migration will be estimated by developing a migration rate for the estimation period (1970-73), based on IRS Form 1040-1040A returns (or exemptions) for each estimate area and applying this rate to an appropriate population base (the 1970 census population of the area plus an allowance for subsequent natural increase). IRS tape records for tax year 1969 will be matched by individual social security number with tape records for tax year 1972.

Of course, these files are handled in a confidential manner. The data tapes are manipulated specifically to develop group tabulations on the basis of changes in residence status of individuals.

The 1972 tax returns are the first to have been modified to provide (in addition to the mailing address) the State, county, incorporated place, and township in which the filer lived at the end of 1972. Approximately 70 percent of persons filing their 1972 tax return gave this full area detail. The 1969 file, of course, will not have this residence detail. In order to permit assignment of each taxpayer to specific place of residence, the 1969 returns and the uncoded 1972 returns will be coded to place of residence either by drawing upon the relationship between place of residence and mailing address in the fully coded 1972 returns or by coding on the basis of 1970 census place geography.

The matching process provides a matrix of place of residence in 1970 by place of residence in 1973 for individual taxpayers identified in the file for those two periods. Thus, for each area we know the number of taxpayers living in the area in both years (non-movers), taxpayers in the area in 1970 but not in 1973 (out-migrants), and those in the area in 1973 but not in 1970 (in-migrants). The number of in-migrants minus out-migrants for each area yields a figure representing net migration of taxpayers. A net migration rate is derived by dividing the net migrants by an appropriate base obtained from the number of cases matched in the file for each area. Problems involved in this migration estimate are discussed in some detail later in this paper.

Migration rates varying from ranges of tolerance set up by size of place will be screened. Where rates are unacceptable, consideration will be given to substituting the migration rate for the next larger territorial subdivision containing the place or for other broader areas within the county. This procedure will also be used to estimate migration rates for new incorporations since 1970.

Institutional and college populations and to a lesser extent Armed Forces are not always reflected in the net migration component derived as above. Many of these people, of course, will not file tax returns. The specific 1970 locations of these populations will be identified by using special 1970 census tabulations of group quarters and Armed Forces population by type for all areas that will be estimated.

Separately we will be compiling current statistics from other sources covering the institutional population, college students, and Armed Forces. Where there are large changes since 1970 in these special populations for a given place and it is determined that they are inadequately covered by the estimating procedure, the population of the place will be adjusted to reflect this change.

These components of population change then are added to the updated population base to produce current population estimates for counties and, independently, for all subunits of counties covered under the Revenue Sharing law together with county remainders. Estimates for the county subunits will be adjusted to the independent estimate at the county level.

Under the Federal-State Cooperative Program a separate series of 1972 and provisional 1973 county estimates for counties will be available for comparison. The Revenue Sharing estimates will be reviewed against the FSCP estimates to verify the reasonableness of the county totals.

In all probability if this new procedure proves effective it will then be averaged with the other procedures used in the Co-op estimates to produce what should be a more accurate county estimates series. Its weight in the averaging will depend on our evaluation of its effectiveness. The estimates for the county subunits will then be adjusted to this "best" county total.

In future years the estimating procedure is intended to provide changes since the last update rather than since the last census. This should take advantage of an expected improvement in the tax return records series.

Tests of Accuracy

Like any new procedure, the population update mechanism described here must be proved effective and reliable. Since this procedure has only been explored since the middle 1960's, testing of its accuracy at the county and place level has necessarily been limited to date. A test covering two dozen counties and moderate-sized cities was conducted for the period 1968 to 1970 using areas with available special censuses or reliable independent estimates to make comparisons with the IRS procedure. This test suggested an average difference for the 2-year period of slightly less than 2 percent using income tax returns, and a somewhat larger difference using tax exemptions. This is an encouragingly small error. Over a longer period of time, however, there is no certainty how the level of error will be affected. The migration detail was obtained using 1970 census coding relationships of street address to place of residence.

For the period 1970 to 1973 we have planned a much more elaborate test of accuracy of the method for some 86 small areas where special censuses were conducted by the Bureau in April and May 1973. Included are counties and places most of which were selected by a random sample of incorporated places and revenue sharing townships under 20,000 population. A handful of areas were included to measure the adequacy of the procedure in estimating military concentrations, college communities, rural areas, and so on. Both population and income were obtained for these areas. An additional 265 special censuses of places of various sizes routinely taken by the Bureau during the period November 1, 1972 to October 31, 1973 are also being included in the test, although no income detail is available for those areas. This test should give us a much better feeling for the adequacy of the approach, especially for small areas.

Some Problems Peculiar to the First Round of Estimates

In developing the population estimates for revenue sharing we are now resolving several problems peculiar to this first round of estimates. First we plan to generate migration rates based on both income tax returns and exemptions claimed in those returns. Exemptions will be adjusted to exclude duplications for blindness and age. Intuitively exemptions would seem to be the more appropriate series, since it should be more closely related to population than returns, which should relate more to households. In what limited testing we have done to date, that for 1968-70, the results slightly favor the use of returns. The 1973 test may provide an indication of the superiority of one series or the other. The large-scale routine processing of estimates for 1973 also could demonstrate that there is little to choose between the two series. If the test favors the use of returns or suggest a stand-off in the applicability of the two alternatives, we will fall back on returns as the simpler series to process and control over time.

In matching the tax return records the obvious intent is to determine change in place of residence. Since no question was asked on place of residence on the 1969 tax return, the 1969 file has only mailing address. How to reconcile the series with a 1972 file essentially reflecting place of revenue sharing residence has been one of the major stumbling blocks in perfecting this estimates system. The Bureau does not have an address system with the capability to code place of residence to the 1969 file on the basis of the 1970 census for most places under 30,000 population with a high degree of certainty. Below that level residence coding would be done by post office name and ZIP code. If this coding system were utilized for 1969, there would always be the question of consistency with 1972. What errors would be introduced by matching data residence-coded by two entirely independent systems in the two return years?

Resolution of problems in relating these 1969 and 1972 IRS migrant files is being handled by a two-pronged approach. First, to take advantage of the full residence coding that exists for 70 percent of the 1972 file, an address coding guide is being prepared from this file relating place of residence with mailing address. The 1969 file and remaining 1972 uncoded records

would be assigned residence codes on the basis of the knowns. This system should prove effective to the extent that the missing 30 percent of the 1972 file doesn't have a distributional bias. In other words, is the 1972 coded file representative of the full file?

Results of this coding procedure will be available early in the processing. If they are discouraging, then we are prepared as a secondary line of action to generate place of residence for the 1969 file on the basis of the best available residence coding from the 1970 census.

We will not be faced with this sort of problem in future years, for place of residence detail will be picked up on income tax returns. This will insure a continuing consistent residence series.

Since a filer cannot be identified as a migrant unless his records can be matched for the two tax return years, you will note that the migration estimates for each revenue sharing area are based on deriving a migration rate in which both the numerator (migrants) and denominator (all returns) are matched returns. This migration rate for matched returns is applied to the population of the area on the base date. This procedure is accurate insofar as the migration rate of the matched returns is a reasonable facsimile of the population it represents. The impact of special groups such as new entrants into the labor force, the elderly, and the poor on the migration rates will vary from area to area. While we would expect the impact of distortions introduced by these groups to be minor for the most part, we will be looking for problem areas where a definite bias may be introduced.

A system for adequate evaluation of the revenue sharing estimates must be set up to function smoothly and quickly. No estimates staff to my knowledge has ever had to face up to a review of 39,000 estimates. We will be required to conduct a review and provide the Treasury Department with a recommendation of the potential utility of the estimates within a period of no more than four months. Obviously much of the review function will have to be handled by computer, with confidence limits established in advance and variant estimates being identified for special review. Of course independent estimates at the county level provided under the Federal-State Cooperative Program will be available to confirm county trends. Below the county level, and particularly for the thousands of very small places, few outside comparisons will be available.

Conclusion

Despite the problems that I have enumerated here about the proposed estimates system, we obviously feel that the procedure has great promise. Among its more obvious advantages for estimating the population (in addition to its simplicity) is the direct use of data in developing the component estimates. Reported births and deaths are used to estimate natural increase, and migration is measured by the change in residence on the tax records. The use of the tax data represents a breakthrough in the use of individual records rather than classes of data in deriving the estimates.

Above all, the technique provides an opportunity for estimating the population of very small areas. Few if any currently used methods can be effectively applied below the county level because of inadequacies in, or lack of, data input. Since up to 36,000 of our revenue sharing areas are areas smaller than counties and about 26,000 of them are below 2,500 population, we now have a potential method for updating areas where for all practical purposes none existed before.

The cost of manipulating 500 computer tapes for each of two tax years in order to match individuals and determine their migration status is substantial. This is a process that only can be done centrally within the Federal government, for others will not have access to the files. The rewards in terms of the hoped-for creation of quality estimates can make the cost worthwhile, considering the amounts to be distributed under the revenue sharing program.

Development of the Components of Personal Income and Money Income for the States and Counties

Edwin J. Coleman, Bureau of Economic Analysis

The purpose of this paper is to touch on some of the methodological aspects of county income estimation and their importance and impact on the quality, quantity, and timing of BEA's annual measures of State and county personal and money income.

It is worth noting at the outset that the development of State and county intercensal estimates of money income called for the modification of the more comprehensive annual personal income totals to eliminate imputed income and income in kind and to make certain other adjustments in several of the money income items. It, therefore, follows that the quality of BEA's estimate of State and county money income depends on the reliability of the basic county personal income series. Exhibit A lists the specific items and dollar values required to make the necessary adjustments and shows the dollar differences between the BEA and Census money income total for 1969.

BEA's local area income estimates are unique in that they provide its users with an economic base profile, that extends over a span of more than 40 years and covers more than 3,000 subareas, which can be combined into any multi-county geographic area needed. That is, it covers the Nation county by county; it is prepared annually; and it is presented in substantial industrial detail.

Because of its conceptual scope and richness of component detail, personal income can be used to measure and track the levels and types of income flows received or earned by individuals residing or working in a specific political subdivision or micro-economy. It has many uses. One of its more important applications include its use as a tool for implementing Revenue Sharing as well as other Federal, State, or local government programs requiring an analysis of economic activities in, or the economic well-being of the residents of a particular area. Some other applications include its use as a measure of the size of the consumer market; as an indirect measure of the industrial market; as a measure of the quality of a market and the economic welfare of an area's residents; as an indicator of an area's economic health; as an analytical framework which can be used in assessing an area's economy and against which alternative developmental programs can be evaluated or as a basis for gauging the capacity of an area to support public service expenditures.

BEA's interest in the measurement and analysis of the income of persons and the subnationalization of the personal income sector of the national income and product accounts goes back a long way. Our descent into the subterranean world of regional income estimation began in the mid-thirties with the construction of a series termed "State Income Payments" to individuals. The Bureau's current effort to measure personal income, at the local-area level, highlights the fact that BEA's regional program in the seventies, as in the thirties, is in keeping with the Bureau's policy of providing the economic measures needed by those dealing with the Nation's economic problems, policies, and programs. My goal in this paper is to provide you with an overview of the county income estimating process so that you may judge for yourself the usefulness of the personal income estimates to your programs or agencies.

County personal income is the sum of several hundred individually estimated component flows, both monetary and nonmonetary and it encompasses most forms of income flowing to persons including the income received by individuals from business establishments, Federal, State, and local governments, household, institutions, and foreign governments.

Several of the more important characteristics of the county income estimates are that they are constructed from a wide variety of statistical information not necessarily designed for income estimation; that this information consists mainly of data obtained from business and government records showing disbursements to individuals; that the BEA national and State estimates of personal income are used as statistical "frames of reference" in which these higher level totals are allocated to the counties on the basis of the best available data; and that a detailed procedure of estimation is followed to maximize reliability and to obtain analytically useful information by county on income by type and by industry. This use of extensive detail is the hallmark of a reliable measurement effort because it reflects with reasonable reliability the interindustry or intercomponent differentials at both the State and county levels and it provides the detail needed to audit the sources and methods of estimation. This detailed measurement technique has an added virtue in that it gives us the opportunity to take maximum advantage of the benevolent "law" of offsetting error.

It should also be emphasized that while the BEA estimates of county personal income and employment incorporate a great deal of basic data which may be presumed, from their general characteristics to be "reliable," it also uses "other" data of differing quality, scope and relevance. To bridge the gaps or deficiencies in these "other" less good statistical source materials, resort must be made to assumptions and indirect procedures.

The statement that the county estimates are subject to error and that the degree of error is difficult to specify in exact terms is, of course, a cliché. In the case of county income the difficulty of establishing the error factor quantitatively is compounded because the amount and quality of basic county data—and hence the extent of reliance on indirect methods and procedures—varies considerably over the 40 year estimating span, and has not had uniform effects on the State and county income totals because of geographic differences in the composition of the income flows. In any case the measurement of error is a generic problem and the inability to quantify error and therefore the degree of reliability, is not confined to the estimates of county or State income. It applies generally to the field of national income accounting, and indeed, encompasses the entire range of economic statistics.

The actual estimation and distribution of county personal income takes place within the framework of the Regional Economics Information System and is the product of what can be described as a generalized dynamic data system made up of four elements: (1) an input data base, (2) an estimating operation, (3) a component aggregation procedure and (4) a retrieval operation. This system can be compared to a manufacturing process. In the first stage, we mine and refine the raw material; in the second stage, we fabricate the component parts; in the third step, the component parts are assembled; and in the fourth step the final product is distributed.

In constructing regional measures of individual income flows, we have tried to live up to BEA's reputation as skilled practitioners of the art of converting data compiled by other agencies for other purposes for use as inputs to our various measurement programs.

As I have indicated, the bulk of the source material used to prepare the county income estimates is not drawn from a coordinated statistical program designed for local measurement, but is culled from administrative records of Federal and State government programs with the remainder of the data coming from the various censuses or from other nongovernment sources.

Several of the more important sources of administrative record information include data generated as the byproduct of the State unemployment insurance programs of the Manpower Administration; the insurance programs of the Social Security Administration; and the Federal tax program of the Treasury Department. Two of the more important census inputs include the survey materials collected as part of the censuses of agriculture and the decennial tabulations of interstate and intercounty commuting flows. The data obtained from these sources yield more than 80 percent of the raw statistical materials needed for the preparation of State and county income estimates.

Any data system based on several million administrative records is subject to a wide range of errors as data are encoded and processed by the agencies administering their respective programs. The user of administrative record files is faced with three alternatives: He may verify each individual record; he can accept the file at face value; or he can develop an edit procedure to catch the major errors in the data base. The first alternative is unrealistic; the second option is irresponsible; and third choice is expensive and timing-consuming. BEA was able to go with the latter option because of the development of several straightforward computerized exception edit routines which have proven to be reasonably effective in pinpointing major errors in the source materials. Without the use of these or similar computerized techniques, our confidence in the reliability of many of the components of BEA's county income series would be greatly reduced and the usefulness of the series as a gauge of economic activity considerably lessened.

Although the use of automated techniques contributes to the quality of the county income estimates, it is not possible or even desirable to mechanize completely either the preparation of the input data base or the estimating procedures because much of the reported statistical information is not directly or wholly suitable for income measurement. Some of the source material has to be refined to adjust for differences in definition. All input data must be edited to identify and correct errors and data gaps must be closed through estimation as part of the step-by-step buildup of aggregate income from the individual component flows.

The estimating techniques used to carry out these tasks require ingenuity and trained judgement on the part of the estimator. The methods used to turn raw data into the component parts of personal income usually do not require the use of sophisticated statistical techniques. They do occasionally involve the use of complex procedures and the estimator does require a finely tuned and responsive computer environment to process the massive amounts of data.

BEA provides the estimator with this environment in the form of an "open-shop" computer facility and a user language designed specifically to handle the types of computations and processes most commonly needed to construct regional economic statistics. This type of EDP environment is essential for the production of reliable county income statistics because of the estimator's need to test and retest and to make frequent changes in methods and techniques with a minimum of reprogramming. Without this somewhat unique balance of professional and machine, we could not carry on the BEA tradition of "handcrafting" the estimates to insure their reliability and conformity to national income accounting concepts and standards. Before I discuss the specific sources and methods of estimation used to conform to these standards, I would like to digress for a moment and comment on the issue of quantity and timing.

One of the most frequent and persistent requests we receive from people involved in regional analysis or planning at the local area level is for more detailed information about county income flows. At the present time, BEA is providing its users with breakdowns of personal income by type of payment and of earnings by broad industrial sector, as shown in exhibit A. The earnings components of personal income, however, are prepared

in considerably more detail than shown in the personal income tables. For example, in order to improve the statistical weighting, estimates of wages are prepared for more than 60 individual industry groups. We cannot, however, publish these detailed wage flows because we are required to comply with more than 50 Federal and State regulations established to protect against the release or publication of information that would reveal the activities of individual establishments.

There is some question in my mind as to the wisdom of applying disclosure regulations in a blanket fashion to such broad measures as wages and employment. However, no changes in the present disclosure laws are anticipated, and we will, of course, continue to comply with the disclosure regulations as they stand. We are attempting to resolve this dilemma by asking our major data suppliers to provide us with information that will permit us to purge the confidential items from our more detailed wages files. But, here, too, it must be recognized that the addition of a disclosure code to a basic data file can be an expensive and timeconsuming procedure and the agencies generating administrative records often exhibit an understandable reluctance to spend money to enhance the statistical value of their data files. Finally, it should be noted that the limits on the release of local area data apply with equal force to all statistics prepared by the Federal Government.

My comments on the problem of timing are of a briefer and more positive nature. Beginning with the 1972 series, BEA's estimates of county personal income and employment will be made available 13 to 15 months after the close of the calendar year to which they refer. This will be 3 to 5 months sooner than in previous years. This earlier release of the county income estimates is the result of a speed-up in the county estimating cycle and reflects BEA's commitment to provide the Census Bureau with State and county estimates of money income on an accelerated time schedule.

This brings me to the central issue—quality and the dry but important task of briefly assessing the statistical accuracy of each of the major types of personal income flows.

In the final analysis, the user must examine the major type of payments which make up the personal income totals and decide for himself whether they are sufficiently reliable for the intended purposes. The major types of income flows, to be briefly discussed, include wages and salaries, various types of supplementary earnings termed other labor income, the net incomes of owners of unincorporated business, property income, government and business transfer payments, and personal contributions to social insurance programs.

The first three types of payments shown in exhibit B, i.e., wages, other labor income (OLI) and proprietors' income, are combined to yield the earnings components shown under total earnings. These earnings are a useful measure for those interested in the functioning of the economy of a given area in that they can answer such questions as: What are an area's industrial sources of income? Does an area contain growing industries with a future potential or is the area characterized by industries which are playing out? Is the present economic structure a balanced one? Also, earnings on a place-of-work basis can serve as an indicator of economic production in an area.

The first type of payment shown in exhibit B is wages and salaries. Total wage and salary disbursements account for about 70 percent of all personal income. The payroll estimates are more complete and reliable than those for any other major type of income payment, and because of their sizeable weight in the total income flows, they impart a large measure of reliability to the estimates of aggregate income at the county level. The bulk of the wage and salary disbursements consist of that cash flow to employees commonly regarded as wages and salaries, inclusive of executives' compensation, commissions, tips and bonuses, and of the value of payments in kind which represent income to the recipient. All disbursements in the current period are covered, including any payments retroactive to past periods. That is, retroactive wages are counted when paid, rather than when earned.

Payrolls are measured before deductions for social security contributions, union dues, or other purposes. The contributions made by employees under the various social insurance programs, although counted in wage and salary disbursements, are not part of the personal income total. They are excluded by means of the deduction, for "personal contributions for social insurance."

Payrolls in the private sector of each subarea's economy cover employees not only of all nonfarm business establishments, but also of farms, private households, hospitals, and private educational, social service and nonprofit institutions. Payrolls in the government sector include those of all Federal, State and local government employees.

Beginning with 1962, the bulk of the county wage and employment data used to estimate payrolls was obtained from the records of the unemployment insurance programs administered by the State agencies of the Manpower Administration.

To prepare a complete measure of wages, it was and still is necessary to make other additions to the Unemployment Insurance (UI) industry data. Several major portions of "covered" industry payrolls still remain outside the scope of the State unemployment insurance laws. These excluded elements consist of wages and salaries of Federally chartered credit unions, Federal Reserve Banks, national banks, (State banks that are members of the Federal Reserve System in New Jersey), electric railways, carrier affiliates in the transportation industry, insurance solicitors on commission basis, and employees' tips. In addition, entire industry groups at one time or the other were not covered by the ES-202 program. These noncovered industries included farms, Federal, State and local government, railroads, private households, hospitals, nonprofit membership organizations, museums, art galleries, etc., private educational services, forestry and fisheries, and a rest-of-world category.

In some instances, payrolls of industrial segments excluded or treated as not covered by the UI program could be estimated quite readily by counties. In others, the task is difficult and the results less satisfactory. Finally, it should be noted that the county wage and salary series for the earlier years have a lesser degree of reliability chiefly because of the absence of a comprehensive body of information comparable to the UI data. Fortunately, the task of constructing a record of wages and salaries for the major industry groups by counties for this earlier

prewar period could be carried out satisfactorily by using data tabulated by the Social Security Administration and the Census Bureau.

To sum up, the statistical derivation of much of this central element of county personal income depends on the availability of county payroll tabulations for those industries in the private and government sector covered by the various State unemployment insurance laws. Principally, though not solely, because of these UI tabulations, the quality of private nonfarm county wages and salaries totals can be rated as excellent.

If you will refer to the exhibit B listing showing the breakdown of personal income, you will note that the earnings component for government is broken down into Federal Civilian, military, and State and local earnings. When we move from an assessment of wages in the private sector to an evaluation of wages in the government sector, the situation becomes somewhat more problematical, because, while the statistical basis for estimating payrolls, in the government sector, has generally been quite good, year to year variations in the relative accuracy of the Federal and State and local segments are more pronounced.

Federal civilian wages and salaries have been covered by the UI program since 1967 and are currently used directly or indirectly to derive county totals for all Federal Civilian wages. These estimates can be considered to be very reliable.

A combined total is shown for State and local government earnings because we have not been able to prepare separate estimates for State or local government employees that were consistent or comparable for all years. County benchmark estimates of State and local government wages and salaries were prepared using Census of Governments tabulations at the SMSA and county level and for the benchmark years are generally considered to be good while estimates for the intervening years were derived by interpolation and are somewhat softer.

The third breakdown of government shows earnings of military personnel. The military component of personal income differs significantly in concept from the measures of gross earnings in the other industrial sectors in that earnings of military personnel stationed in each State or county excludes those amounts withheld by the government and sent to dependents of military personnel.

Military disbursements by local area are thus derived as the sum of two separate flows. The first flow includes the gross earnings of military personnel stationed in each State or county while the second flow nets out the amounts withheld by the Government and sent to their dependents or other individuals in the form of voluntary allotments of pay of military personnel wherever stationed. A noteworthy aspect of this item is that it represents the only element of wage and salary disbursements not received by individuals in an employee status.

Apart from dependency allotments (for which some primary State and county data are available) little direct data on military payroll disbursements are available, and as a result, county estimates of military payrolls are constructed from military strength data obtained from the various service arms of the

military establishment. The lack of payroll data by county is a significant limitation. It has, however, been offset to some degree by the branch-by-branch distribution of State totals and, on the whole, I would rate the county estimates of military payrolls as satisfactory.

There is one additional aspect to the wage estimates worth mentioning, and that is imputed wages. The wage and salary estimates include allowances for the food, clothing, and lodging paid in kind to employees. The concept of valuation for imputed wages is cost to the employer.

As might be supposed, the area of wage imputation is rather imprecise and involves a number of difficult decisions which can only be settled in a pragmatic fashion. For instance, the imputation is confined to food, clothing, and lodging because other types of perquisites, such as medical and recreational services, are generally less important and cannot be estimated satisfactorily from available data. It is frequently difficult, moreover, to determine whether or not a particular type of payment in kind clearly represents an addition to cash wages and salaries. The relative importance of imputed wages varies from industry to industry. Payments in kind covering the cost value of the food and clothing provided enlisted personnel as part of their total pay and allowances are a significant element of military wages.

As to other industrial segments of the State and county estimates, wages in kind (comprising either food or food plus lodging) are of some significance in eating and drinking places, farming, private households, water transportation, hotels, and hospitals. They are quite minor, however, in other areas of private employment.

The second type of payment shown in exhibit B consists of supplementary forms of labor income paid out or accruing in the current period and includes such items as employer contributions under private pension, health and welfare and group insurance plans which accounts for three fourths of all other labor income. Such funds are classified as persons in the national accounts and their income included in personal income. In the county series the "residence" of such funds is taken to be the same as the place of work of the employees on whose behalf the employer contributions are made. This treatment is carried out statistically by allocating to the counties the State totals of employer contributions, industry by industry, on the basis of payrolls or employment. The allocations have been carried out in considerable industry detail, because the ratio of employer contributions to wages and salaries and employment differs widely by industries. The final county distributions, although based on secondary allocators, yield reliable county estimates of other labor income.

The third type of payment is proprietors' income. Proprietors' income measures the net business earnings of owners of unincorporated enterprises. Farmers, independent professional practitioners such as physicians, dentists, and lawyers, and others in a self-employment status fall within the scope of proprietors' income. This income component may be subdivided with respect to source material and methods of estimation into nonfarm and farm proprietors' income.

Estimates of nonfarm proprietors' income which account for 6 percent of personal income nationally are based on data collected by the Treasury Department, the Manpower Administration, the Census Bureau, and various professional associations and are less good and more difficult to estimate than either wages or other labor income. Nevertheless, based on comparisons of our present estimates with recent IRS based benchmarks, I would rate the current county measures of total nonfarm proprietors' income as good.

The situation with regard to the income of farm proprietors is somewhat similar. Estimates of the net income of farm proprietors at both the State and county level are equal to, and derived statistically as, the gross income of farmers minus production expenses, and as such, represents one of the most difficult components of income to estimate.

The concepts underlying the BEA county estimates of farm income are the same as those used for the National and State farm income estimates prepared by the U.S. Department of Agriculture. The major conceptual difference between the two series is that the USDA totals include, and the BEA totals exclude income of corporate farms.

The total net farm income estimates measure income arising out of the current year's production in the farm sector. In order to arrive at this level, income is adjusted by the value of the net change during the year in farm inventories of livestock and crops held for sale.

The methods used to generate farm proprietors' income rely heavily on quinquennial data obtained from the Census of Agriculture and on selected intercensal data prepared by the Economic Research Service (ERS) at the State level and by the Statistical Reporting Service (SRS) at the county level. These data are used, where possible, to bridge the intercensal gap. In addition to these basic inputs, other statistical sources developed in the Department of Agriculture are utilized in the preparation of a fairly detailed income-and-expense statement covering all farms in each State and county. Obviously, the measurement of as volatile a component as farm income is difficult and one should view the farm detail as approximations. Despite their limitations, they represent an extensive and serious effort to measure income flows in the agriculturally oriented rural counties.

The fourth and fifth types of payments, property income and transfer payments can only be measured on a place-of-residence basis since the capital which generates the property income received by the residents of an area may be located anywhere in the world. Similarly, the work for which some transfer payments are received may have been performed in any county in the Nation, or even abroad.

Property income, accounts for about 13 percent of U.S. personal income. It covers the return on capital invested outside of one's own business, and is received in the form of interest, dividends, net rents and royalties. Property income also includes an estimate of imputed rent and interest. The major property income imputation, imputed rent, measures the net rental income accruing to nonfarm residents of owner-occupied

nonfarm dwellings less the normal expenses incurred in home ownership. A similar imputation for farm dwellings is implied in the estimates of the net income of farm operators.

Until recently, the amount of direct data on personal receipts of property income, at the county level, has been minimal. The bulk of the available source materials were indirect, incomplete or summary in nature and as a result, it was not possible to make use of a detailed method of estimation. Instead we had to fall back on a relatively summary allocation procedure based on partial data and the use of several regression equations based on nonproperty income. A more detailed procedure would have improved the statistical weighting and minimized error in that we could have made a better match of the incomplete or indirect data used for allocation and the income flows to be allocated.

The county property income estimates are, therefore, of a lower order of reliability than most other types of county income flows and are subject to appreciable percentage error in both general level and year-to-year movement. This generalization applies particularly to rental income and interest and to a lesser degree to dividend income. We are in the process of upgrading this type of income flow through the introduction of IRS county benchmark totals of dividends and interest. At the present time, we are trying to incorporate the more reliable IRS-based county monetary income items as part of the 1972 estimates.

The fifth type of payment is transfer payments. Transfer payments comprise 10 percent of national personal income. The transfer payments category comprises, in general, receipts of persons from government and business (other than government interest) for which no services are rendered currently. The estimates for this important type of payment are quite good.

Government transfers consist of Federal, State and local government payments to individuals not in return for current services and to private nonprofit institutions such as hospitals and charitable and welfare organizations. This first category includes such items as old-age employee retirement systems, direct relief, and payments to former members of the military establishment.

Business transfers consist of distributions of business output to persons for which no services are received. Included under this heading are such items as individual's bad debts to business, corporate gifts to private nonprofit institutions, cash prizes, and personal injury payments by business other than to employees.

In some instances, transfer payments represent payments for services rendered at some previous time, as in the case of retirement and veterans payments. In other cases, they represent simply a redistribution of income from one group of people to another, such as welfare or relief payments.

The estimates of total transfer payments represent the summation of approximately 45 separate series. Some were obtained through a process of detailed data collection. Others were estimates by means of allocators which vary considerably, both in directness and relevancy.

Currently, directly reported data underlie the estimates of individual items that in combination account for more than two-thirds of total transfers nationally, although the proportion varies from State to State.

In general, the transfer payment components are based on reports of disbursements obtained from the fiscal records of governmental agencies administering the program. In those instances where administrative records were not available use was made of indirect allocators. An example of a secondary allocator is the county distributions of veterans of World War II which were used to apportion several types of veterans payments. Transfer payments components for which the statistical basis is weak, comprise only a small part of total transfers and an almost negligible fraction of total personal income.

Personal contributions for social insurance is the sixth and last type of payment and are to be considered quite reliable. Contributions made by individuals under the various social insurance programs are excluded from personal income by handling them as an explicit deduction item. Payments by both employees and self-employed are included in the series.

The general procedure used to estimate personal contributions to social insurance is to allocate State totals to the counties on the basis of payrolls or proprietors' income in the relevant category of income.

All personal income can be classified under one of the six types of income I have just described. However, to complete the estimating process, and provide a measure of total personal income, an explicit adjustment is necessary to convert those income components measured on a place-of-work basis to a residence basis. Components subject to this situs adjustment includes major elements of wages, other labor income and contributions to social insurance.

These residence adjusted personal income totals, when divided by the total population, yield the estimate of per capita income shown on the computer listing of exhibit C which also presents per capita relatives. The chief uses of per capita income are as a measure of economic welfare and as an indicator of the quality of the consumer market.

One can make the valid argument that per capita income cannot precisely measure real differences in economic welfare among areas without adjustment to take into account the differences in cost of living. However, this is extremely difficult to do. I prefer to make the assumption that the wide variations in per capita income shown in exhibit C exceed any differences

that may exist in cost of living, thereby indicating that there are significant differences in economic welfare among local areas.

Another limitation generally ascribed to per capita income as a measure of welfare is that it is a simple arithmetic average that fails to show the degree of variation in income around the average. That is, critics of the per capita measure point out that of two counties with identical per capita incomes, the population of one may consist of a group of wealthy families and a group of very poor ones, with very few in the middle income brackets. In the second county, with a per capita income the same as the first, the bulk of the families may fall within a comparatively narrow range around the average figure, with a few high income families and a few low income ones. Obviously, the economic status of the residents of these two counties would be quite different, and per capita income would be a poor indicator of economic welfare.

However, the situation described in the first county, though theoretically possible and often used illustratively, is rare or perhaps even nonexistent. That described for the second county approaches the usual. To the extent that the families in a county are distributed around the county average in a normal manner, per capita income forms a satisfactory measure of economic welfare.

A simple solution to the problem would be to provide a per capita figure and an income distribution for every county. This would solve the problem partially, but it is prohibitively expensive to prepare size distributions of income at the county level each year, so we must wait for the periodic decennial census of population. These 10-year benchmarks provide a satisfactory measure with which to evaluate the validity of the per capita figure with respect to dispersion around the average.

The per capita income totals when coupled with the component detail of personal income provides a valuable economic measure for a wide range of users. It is valuable to the businessman who want to know the size, type, and location of markets in a region. It is essential to the legislator or government administrator who needs to know the fiscal capacity, the welfare needs, or the economic health and problems of the jurisdiction for which he is responsible. It is of use to the researcher interested in the functioning of an area's economy, and to the teacher—high school or university—whose area of instruction covers the economic base on which the livelihoods of the residents depends. It is a requirement for the development planner interested in improving an area's economic base and it is important to the man-on-the-street whose economic well-being depends on the strength of each local area personal income base.

Exhibit A

Reconciliation of 1969 Total Personal Income With Census Money Income

(\$000,000)

	Income	BEA	CENSUS
I. Wages and Salaries (Personal Income)		505,204	
Less: Imputed income		2,032	
Military family allotments		637	
Plus: Military Reserve Pay (excluding pay-in-kind)		1,148	
Directors' fees		443	
Selected government "Other Labor Income"		83	
Equals: Wages and Salaries (Census-concept Money Income)		504,209	500,140
II. Nonfarm Proprietors' Income (Personal Income)		50,450	
Less: Income retained by fiduciaries		284	
Other noncash income (inventory valuation adjustment (IVA) bad debts, home-owner construction		-210	
Patronage refunds of farmer's co-ops		444	
Equals: Nonfarm proprietors' income (Census-concept Money Income)		49,932	47,917
III. Farm Proprietors' Income (Personal Income)		16,741	
Less: Imputed income		1,282	
Farm land rent paid to farmers		571	
Change in inventory		65	
Plus: Patronage refunds of farmers' co-ops		444	
Equals: Farm proprietors' income (Census-concept Money Income)		15,267	8,836
IV. All Other Income			
1. Property income (Personal Income)		106,147	
Less: Income retained or spent by institutions on behalf of individuals		7,772	
Imputed income		37,058	
Excess of accrued interest over interest paid on U.S. Savings Bonds		674	
Plus: Farm land rent paid to farmers		571	
Equals: Property income (Census-concept Money Income)		61,214	

Exhibit A--Continued

2. Transfer Payments (Personal Income)	65,768	
Less: Business and government payments to nonprofit institutions	3,134	
Lump sum benefits payments	3,218	
Income received, but not spendable without restriction	7,203	
Other noncash income items	1,156	
Plus: Military family allotments	637	
Equals: Transfer payments (Census-concept Money Income)	51,694	
3. Other Labor Income (Personal Income)	28,415	
Less: Workmen's compensation hospital and medical benefits	920	
Employer contributions to private pension, health and welfare funds	23,645	
Military Reserve Pay (including pay- in-kind		
Directors' fees	443	
Selected government OLI	83	
Equals: Other Labor Income (Census-concept Money Income)	2,084	
All other income (Census-concept Money Income)	121,940	81,041
Property plus transfer payments plus other labor income	114,992	
Estimated income from private pensions and annuities	6,948	
Total money income: Census-concept	691,348	637,934

Exhibit B

(300-992-415-000) Madison, Wisconsin SMSA

August 9, 1973

Personal Income by Major Sources and Earnings by Broad Industrial Sector

Income and Earnings	1967	1968	1969	1970	1971
(Thousands of dollars)					
Total personal income.....	927,021	1,022,347	1,124,609	1,245,808	1,338,057
Total wage and salary disbursements ¹	643,695	719,852	811,155	897,993	959,684
Other labor income.....	22,799	25,405	30,545	35,217	39,960
Proprietors income.....	83,506	89,010	86,117	86,409	93,886
Farm proprietors income.....	22,990	27,163	23,443	26,526	28,344
Nonfarm proprietors income.....	60,516	61,847	62,674	59,883	65,542
Property income.....	147,386	151,285	158,566	177,201	185,610
Transfer payments.....	59,050	68,512	75,521	90,005	104,534
Less: Personal contributions for social insurance.....	29,415	31,717	37,295	41,017	45,617
Total earnings ²	750,000	834,267	927,817	1,019,619	1,093,530
Farm earnings.....	25,197	29,470	25,996	29,355	31,410
Total nonfarm earnings.....	724,803	804,797	901,821	990,264	1,062,120
Government earnings.....	234,986	263,480	295,541	341,992	370,445
Total federal.....	36,485	31,817	34,214	39,051	42,787
Federal civilian.....	25,634	27,227	29,917	34,608	38,079
Military.....	10,851	4,590	4,297	4,443	4,708
State and local.....	198,501	231,663	261,327	302,941	327,658
Private nonfarm earnings.....	489,817	541,317	606,280	648,272	691,675
Manufacturing.....	126,780	134,281	153,744	164,453	165,660
Mining.....	1,330	1,448	1,683	1,721	1,958
Contract construction.....	58,660	71,904	80,942	83,168	90,333
Transportation, communication, and public utilities.....	38,086	40,936	47,534	50,117	52,095
Wholesale and retail trade.....	119,180	129,896	141,402	151,187	164,145
Finance, insurance, and real estate.....	43,152	47,743	52,746	56,342	62,320
Services.....	100,320	111,883	124,272	136,653	149,742
Other.....	2,309	3,226	3,957	4,631	5,422
(Percent of U.S.)					
Total personal income.....	.1482	.1493	.1507	.1554	.1561
Total wage and salary disbursements ¹1535	.1563	.1606	.1672	.1688
Other labor income.....	.1024	.1002	.1075	.1097	.1094
Proprietors income.....	.1344	.1386	.1282	.1294	.1342
Farm proprietors income.....	.1550	.1850	.1400	.1572	.1636
Nonfarm proprietors income.....	.1279	.1249	.1242	.1200	.1245
Property income.....	.1629	.1549	.1494	.1555	.1552
Transfer payments.....	.1140	.1150	.1148	.1133	.1117
Less: Personal contributions for social insurance.....	.1439	.1396	.1419	.1469	.1466
Total earnings ²1489	.1516	.1544	.1603	.1620
Farm earnings.....	.1432	.1674	.1307	.1441	.1497
Total nonfarm earnings.....	.1491	.1511	.1552	.1608	.1624
Government earnings.....	.2813	.2827	.2913	.3044	.3048
Total federal.....	.1012	.0800	.0806	.0854	.0889
Federal civilian.....	.1178	.1132	.1163	.1210	.1241
Military.....	.0760	.0292	.0257	.0259	.0270
State and local.....	.4179	.4335	.4427	.4550	.4463
Private nonfarm earnings.....	.1217	.1232	.1265	.1288	.1299
Manufacturing.....	.0856	.0831	.0880	.0930	.0917
Mining.....	.1249	.0259	.0288	.0269	.0298
Contract construction.....	.1966	.2186	.2182	.2134	.2141
Transportation, communication, and public utilities.....	.1086	.1077	.1140	.1110	.1065
Wholesale and retail trade.....	.1421	.1428	.1433	.1436	.1460
Finance, insurance, and real estate.....	.1674	.1652	.1692	.1725	.1718
Services.....	.1372	.1400	.1404	.1420	.1446
Other.....	.1584	.2065	.2204	.2514	.2759
(Percent of total personal income)					
Total personal income.....	100.00	100.00	100.00	100.00	100.00
Total wage and salary disbursements ¹	69.44	70.41	72.13	72.08	71.72
Other labor income.....	2.46	2.48	2.72	2.83	2.99
Proprietors income.....	9.01	8.71	7.66	6.94	7.02
Farm proprietors income.....	2.48	2.66	2.08	2.13	2.12
Nonfarm proprietors income.....	6.53	6.05	5.57	4.81	4.90
Property income.....	15.90	14.80	14.10	14.22	13.87
Transfer payments.....	6.37	6.70	6.72	7.22	7.81
Less: Personal contributions for social insurance.....	3.17	3.10	3.32	3.29	3.41
Total earnings ²	80.90	81.60	82.50	81.84	81.73
Farm earnings.....	2.72	2.88	2.31	2.36	2.35
Total nonfarm earnings.....	78.19	78.72	80.19	79.49	79.38
Government earnings.....	25.35	25.77	26.28	27.45	27.69
Total federal.....	3.94	3.11	3.04	3.13	3.20
Federal civilian.....	2.77	2.66	2.66	2.78	2.85
Military.....	1.17	.45	.38	.36	.35
State and local.....	21.41	22.66	23.24	24.32	24.49
Private nonfarm earnings.....	52.84	52.95	53.91	52.04	51.69
Manufacturing.....	13.68	13.13	13.67	13.20	12.38
Mining.....	.14	.14	.15	.14	.15
Contract construction.....	6.33	7.03	7.20	6.68	6.75
Transportation, communication, and public utilities.....	4.11	4.00	4.23	4.02	3.89
Wholesale and retail trade.....	12.86	12.71	12.57	12.14	12.27
Finance, insurance, and real estate.....	4.65	4.67	4.69	4.52	4.66
Services.....	10.82	10.94	11.05	10.97	11.19
Other.....	.25	.32	.35	.37	.41

¹Primary source for private nonfarm wages: Wisconsin Dept. of Ind., Labor and Human Rel.²Earnings is the sum of wages, other labor income and proprietors' income.

Exhibit B—Continued

(300-992-415-000) Madison, Wisconsin SMSA

August 9, 1973

Personal Income by Major Sources and Earnings by Broad Industrial Sector

Income and earnings	1967	1968	1969	1970	1971
	(Percent of total earnings)				
Total earnings.....	100.00	100.00	100.00	100.00	100.00
Farm earnings.....	3.36	3.53	2.80	2.88	2.87
Total nonfarm earnings.....	96.64	96.47	97.20	97.12	97.13
Government earnings.....	31.33	31.58	31.85	33.54	33.88
Total federal.....	4.86	3.81	3.69	3.83	3.91
Federal civilian.....	3.42	3.26	3.22	3.39	3.48
Military.....	1.45	.55	.46	.44	.43
State and local.....	26.47	27.77	28.17	29.71	29.96
Private nonfarm earnings.....	65.31	64.89	65.34	63.58	63.25
Manufacturing.....	16.90	16.10	16.57	16.13	15.15
Mining.....	.18	.17	.18	.17	.18
Contract construction.....	7.82	8.62	8.72	8.16	8.26
Transportation, communication, and public utilities.....	5.08	4.91	5.12	4.92	4.76
Wholesale and retail trade.....	15.89	15.57	15.24	14.83	15.01
Finance, insurance, and real estate.....	5.75	5.72	5.68	5.53	5.70
Services.....	13.38	13.41	13.39	13.40	13.69
Other.....	.31	.39	.43	.45	.50
(Location quotient for total earnings) ¹					
Farm earnings.....	.9628	1.1031	.8459	.9000	.9228
Total nonfarm earnings.....	1.0013	.9966	1.0053	1.0033	1.0025
Government earnings.....	1.8896	1.8642	1.8857	1.8992	1.8822
Total federal.....	.6788	.5277	.5227	.5327	.5484
Federal civilian.....	.7917	.7460	.7523	.7533	.7648
Military.....	.5106	.1930	.1655	.1630	.1667
State and local.....	2.8070	2.8599	2.8657	2.8316	2.7562
Private nonfarm earnings.....	.8172	.8124	.8188	.8034	.8017
Manufacturing.....	.5746	.5482	.5700	.5804	.5659
Mining.....	.1698	.1667	.1856	.1700	.1856
Contract construction.....	1.3209	1.4415	1.4133	1.3312	1.3216
Transportation, communication, and public utilities.....	.7299	.7106	.7378	.6930	.6575
Wholesale and retail trade.....	.9544	.9419	.9276	.8961	.9015
Finance, insurance, and real estate.....	1.1230	1.0895	1.0944	1.0759	1.0615
Services.....	.9221	.9229	.9090	.8857	.8924
Other.....	1.0690	1.3929	1.4333	1.5517	1.7241

¹Earnings is the sum of wages, other labor income and proprietors' income.

Exhibit C

Per Capita Personal Income in Dollars and Percent of National Average by SMSA's and Non-SMSA Counties in Selected Years, 1950-72

(Residence Adjusted)

SMSA or County	Dollars										Percent of national average									
	1950	1959	1962	1965	1967	1968	1969	1970	1971	1972	1950	1959	1962	1965	1967	1968	1969	1970	1971	1972
MARYLAND																				
SMSA's																				
Balt., Md.....	1,686	2,272	2,526	2,860	3,270	3,541	3,802	4,060	4,277	4,588	113	105	107	103	103	103	103	103	103	102
Wash., D.C.-Md.-Va....	2,054	2,715	3,031	3,556	3,938	4,274	4,584	5,029	5,389	5,862	138	126	128	129	124	124	124	124	129	130
Wilm., Del.-N.J.-Md..	2,114	2,786	3,043	3,531	3,865	4,134	4,465	4,678	4,996	5,346	142	129	128	128	122	120	120	119	120	119
Non-SMSA counties																				
Allegany.....	1,186	1,746	2,021	2,296	2,630	2,918	3,129	3,426	3,651	3,885	79	81	85	83	83	85	84	87	88	86
Calvert ¹	894	1,310	1,572	1,887	2,102	2,364	2,639	3,179	3,840	4,175	60	61	66	68	66	69	71	81	92	93
Caroline ¹	1,283	1,681	1,974	2,356	2,610	2,966	3,474	3,662	3,591	3,823	86	78	83	85	82	86	94	93	86	85
Dorchester.....	1,026	1,463	1,817	2,148	2,568	2,821	3,219	3,593	3,702	4,106	69	68	77	78	81	82	87	91	89	91
Frederick ¹	1,092	1,675	1,989	2,430	2,866	3,168	3,491	3,768	3,933	4,275	73	78	84	88	90	92	94	96	94	95
Garrett.....	655	1,077	1,256	1,524	1,737	1,918	2,126	2,372	2,596	2,943	44	50	53	55	55	56	57	60	62	66
Kent.....	934	1,516	1,966	2,471	2,992	3,202	3,533	3,743	3,728	4,079	63	70	83	89	94	93	95	95	90	91
Queen Annes ¹	859	1,280	1,595	2,051	2,450	2,684	3,065	3,423	3,429	3,641	58	59	67	74	77	78	83	87	82	81
St. Marys.....	1,438	1,678	1,807	2,003	2,454	2,769	2,975	3,440	3,548	3,951	96	78	76	72	77	81	80	87	85	88
Somerset.....	791	1,340	1,595	1,784	2,019	2,452	2,859	2,951	3,064	3,449	53	62	67	64	64	71	77	75	74	77
Talbot.....	1,241	1,745	2,216	2,665	3,335	3,764	4,147	4,535	4,691	5,090	83	81	94	96	105	110	112	115	113	113
Washington.....	1,405	2,162	2,221	2,586	3,089	3,336	3,563	3,703	3,935	4,313	94	100	94	93	97	97	96	94	95	96
Wicomico.....	1,377	1,752	2,107	2,395	2,751	3,185	3,635	3,933	3,979	4,292	92	81	89	87	87	93	98	100	96	96
Worcester.....	1,291	1,600	1,924	2,313	2,612	2,937	3,497	3,728	3,796	4,228	86	74	81	84	82	85	94	95	91	94

¹This county has a net residence adjustment to total personal income of 15 percent or more in one or more years.

Preparation of Final Revenue Sharing Estimates of Money Income for Political Jurisdictions

Roger A. Herriot
Bureau of the Census

I. Introduction

The current Revenue Sharing program requires a periodic updating of the data elements used to allocate the funds. Current plans call for an update of all data elements, including income, for all of the 38,000 or so participating governmental units. If successful, these updates will bring the population estimates up to mid-year 1973 and will produce income estimates for calendar year 1972.

These estimates represent a significant expansion of the Census Bureau's Administrative Record Estimation Program. For many years the Census Bureau has used data from administrative sources to make intercensal population estimates for States and counties. This, however, is the first time that administrative data will be used to update the Census Bureau income estimates. In addition, it is the first time that such estimates will be prepared for subcounty governmental units. Although the impetus for this extension was the Revenue Sharing program, it is hoped that these estimates will prove useful for many other purposes as well.

This paper is divided into two major sections, the first section will discuss the sources of the data to be used in developing the updated estimates. The data will be examined with regard to their conceptual and geographic dimensions. The final section will present the methods to be used in constructing the estimates for each governmental level. It should be noted that the procedures are still experimental and the particular techniques presented may not be the ones which are finally used. In many cases it will not be possible to decide the final procedures until the necessary empirical data are available.

II. Nature and Sources of the Data

A. The 1970 Census

For the purpose of updating the population and income estimates, special tabulations from the 1970 census 20-percent sample tapes were prepared. For income estimating purposes, three sets of data were tallied. The first set consists of special "journey-to-work" statistics prepared for the Bureau of Economic Analysis. These statistics are used to transform wage and salary estimates from a where-earned to a where-received basis. The second set of statistics is used in making the "death

adjustment" which will be described later. For each governmental unit these statistics include the number of persons receiving and the mean amount of each type of income received by age and sex. In the final set, counts of special population universes, such as the number of veterans, the number of student veterans receiving "other income", the number of persons unemployed, etc., were prepared for use in allocating various income types to subcounty governmental units. It should be noted that for subcounty governmental units, which had qualifying annexations or which incorporated since the 1970 census was taken, the above statistics reflect their status as of December 31, 1972.

B. The Bureau of Economic Analysis Data

For many years now the Regional Economics Division of the Bureau of Economic Analysis (BEA) has prepared annual estimates of Personal Income for States, SMSA's, and non-SMSA counties. They have developed an elaborate system for securing, on a timely basis, the administrative data upon which their estimates are based. In order to avoid needless duplication of effort and cost and to take full advantage of existing governmental expertise, BEA agreed to make the necessary adjustments to their Personal Income estimates and prepare estimates of money income by type according to Census Bureau definitions for States and counties. The adjustments needed and the procedures used to develop these estimates are discussed in a companion paper by Edwin Coleman.

Data constraints however do not allow BEA to make estimates for independent cities such as those which exist in Virginia and some other States, or to make separate estimates for the boroughs of New York City. We will develop estimates for such places using the subcounty estimation methods to be discussed later.

C. Internal Revenue Service Data

1. Data Available

The major data set for updating income estimates for governmental units below the county level are extracts of data from the 1969 and 1972 Individual Master File (IMF). These extracts contain certain items of information from the income tax return of all taxpayers. For income estimation purposes the relevant items of data included on these files are as follows: for

1969—Adjusted Gross Income (AGI), Wages and Salaries, Interest and Dividends in AGI combined, an indicator as to whether or not a Schedule F (Farm Proprietors Income) was filed, and the number of unduplicated exemptions (number of exemptions excluding the additional exemptions for being 65 years or blind); for 1972—Adjusted Gross Income, Wages and Salaries, Interest and Dividends in AGI, indicators as to whether or not Schedule C (Nonfarm Proprietors income), Schedule D (Capital Gains), Schedule E (Pension, Rent, Royalties, Estates, and Trusts, and Partnership income) and Schedule F (Farm Proprietors income) were filed, the number of unduplicated exemptions, and a geographic code. Needless to say once this information is received by the Census Bureau, it becomes subject to the same regulations and laws concerning disclosure that are in force for Census Bureau information.

2. CONCEPTUAL DIFFERENCES BETWEEN IRS AND CENSUS

To be useful in the income updating procedure, the IRS concept and coverage must be at least reasonably comparable to the Census Bureau concept and coverage for a given number type. Two general coverage problems exist. Whereas the Census Bureau excludes, by definition, the income of persons less than 14 years of age and persons who died before the census enumeration, this income would be reported to IRS if the amounts were large enough to require filing. On the other hand, persons with small amounts of adjusted gross income need not file a tax return and thus this income is not counted in the IRS data. This coverage problem is minimal for wage and salary recipients, however, because most of these people probably file returns in order to get their withholdings back. The under-coverage appears to be somewhat more important for interest and dividends however, particularly among the elderly who have higher cut-off requirements for filing and who have a much lower probability of having wages and salaries. For dividends there is also a conceptual difference in that up to \$100 of dividends per return (\$200 for joint returns) is excluded from the IRS data.

The income reported to IRS on schedules C and E (C/E income) is roughly comparable to the sum of nonfarm self-employment income, net rental income, estates and trust income, royalties, and income from public and private pensions as defined by the Census Bureau. Some important coverage differences however do exist. For example, only part of the pension income reported to IRS is taxable and therefore included in AGI. On the other hand, income from farm partnerships and small business corporations is included on Schedule E income. Even with these differences, we believe that a useful allocator can be constructed to allocate the above income types using IRS C/E income.

This allocator will be constructed by imputing amounts of C/E income to returns on the 1972 IRS file which have schedule C or E indicators. These amounts will be developed from the 1971 IRS Statistics of Income sample file. At present, regression and other techniques to impute amounts are being explored, but we will need to look at some empirical results before settling on the best method.

3. GEOCODING OF IRS DATA

In order to improve the usefulness of the IRS data for updating population and income estimates, a set of questions

concerning the location of the taxpayer's residence was included on the 1972 tax returns. On the basis of this information, IRS was to determine a 12-digit geographic code. The taxpayers response to these questions was not, however, overwhelming. Complete geographic information was supplied on only 75 percent of the returns. The remaining 25 percent, about 18 million returns, were coded by IRS to State only on the basis of the information furnished or from the mailing address.

The Census Bureau plans to assign complete geographic codes to the partially coded returns. This will be done using the Bureau's Geographic Base Files or a "coding guide" developed from the 1972 IRS file. If it works, the coding guide procedure has some advantages over the Geographic Base Files in that it can provide geographic codes for all minor civil divisions (MCD's). The coding guide procedure is discussed in the appendix of this paper.

The 1969 IRS file will be geocoded by matching a taxpayers 1969 and 1972 tax returns. If the taxpayer didn't move during that period and he had a complete geographic code on his 1972 return, the 1972 geographic code will be transferred to the 1969 return. If there is no matching 1972 return, or if the taxpayer had moved or had an incomplete 1972 geographic code, the 1969 return will be geocoded using the same technique as used to code the 1972 tax return with incomplete geographic codes.

D. Other Administrative Data

For the two largest sources of transfer payments, i.e., Social Security and Veterans Administration payments, we were able to secure data on the number of recipients and the amounts received by 5-digit zip code. As zip code delivery areas do not conform very well to governmental unit boundaries, these data will be useful only for fairly large cities where most of the people living in the city's zip code areas also live inside the city boundaries. The methods we have developed for using this data in the updating procedures will be discussed in the following section on methods.

III. Methods of Estimation

A. State and County Estimates

Estimates for six classes or types of income will be developed for each governmental unit. These income types are as follows: Wages and salaries (W&S), earnings from nonfarm self-employment (NFSE), earnings from farm self-employment (FSE), Social Security payments (SS), public assistance payments (PA), and all other sources (OY). At the State level, 1972 estimates for each of the above types will be developed by applying the percent change (1969 to 1972) in the BEA State amounts to the 1969 census State aggregate amount.

Except for farm self-employment, 1972 estimates for each type of income for each county will be made in the same manner as those at the State level. Because aggregate farm self-employment income is sometimes negative at the county level, the use of percentage change adjustments will not always produce the appropriate adjustment. For this income type at the county level, the 1972 census estimate will probably be

computed by adding the actual dollar changes in the county farm self-employment income from 1969 to 1972 as estimates by BEA to the 1970 census aggregate amount. The county estimates for each income type will then be raked up to the 1972 State estimates.

B. Subcounty Estimates

Because of the uncertainties involved in imputing geocodes to some 18 million returns, two methods will be used to develop estimates of wages and salaries for cities, towns, townships, etc. The first method is to assume that the

subcounty distribution of aggregate IRS wages and salaries is correct and allocate the county wage and salary control total to subcounty governmental units according to the distribution of IRS wages and salaries within the county. The second procedure is to assume that the 1970 census wage and salary amount per capita increased by the same proportion as the IRS wage and salary amount per exemption from 1969 to 1972. The 1969 amount per exemption will be computed using 1972 filing requirements to compensate for the substantial changes in filing requirements from 1969 to 1972. Both procedures have their advantages and disadvantages but the final choice as to which to use will have to await empirical results.

The farm self-employment estimates as well as those for public assistance will be made by assuming that the relative relationship of the 1970 census amounts per capita for the places to that of the county remains constant over time. The procedures for making the wage and salary, public assistance, and farm income estimates is shown in the following equations:

$$\text{EQ.P1} \quad [72 \text{ CEN } (Y)]_P = \left[\frac{[72 \text{ CEN } A(Y)]_P}{\sum_P [72 \text{ CEN } A(Y)]_P} \right] \times [72 \text{ CEN } (Y)]_C$$

$$\text{EQ.P2} \quad [72 \text{ CEN } A(Y)]_P = [72 A(Y)]_P \times [\text{DEATH ADJ. } (Y)]_P$$

$$\text{EQ.P3} \quad [72 \text{ CEN } A(Y)]_P = \left[\frac{[72 A(Y)]}{[72 \text{ EXEMPT}]} \right]_P \times \left[\frac{[69 \text{ CEN } (Y)]}{[69 \text{ POP}]} \right]_P \times [72 \text{ POP}]_P \times \left[\frac{[69 A(Y)]}{[69 \text{ EXEMPT}]} \right]_P \times [\text{DEATH ADJ. } (Y)]_P$$

$$\text{EQ.P4} \quad [72 \text{ CEN } A(Y)]_P = \left[\frac{[69 A(Y)]}{[69 \text{ POP}]} \right]_P \times \left[\frac{[72 \text{ CEN } (Y)]}{[72 \text{ POP}]} \right]_C \times [72 \text{ POP}]_P \times \left[\frac{[69 A(Y)]}{[69 \text{ POP}]} \right]_C$$

where:

$[CEN (Y)] =$	METHOD =	$[A(Y)]$	EQ. to EST. $[CEN A(Y)] =$
W/S	1	IRS W&S	P2
W/S	2	IRS W&S	P3
FSE	1	CEN FSE of Farmers	P4
PA	1	CEN PA	P4

and where:

P = subcounty part, i.e., cities, townships, (city X township) pieces, etc.

C = County

EXEMPT = No. of exemptions

POP = Population

A death adjustment is included to correct for the fact that some of the income received during 1972 would not have been reported in a census in 1973 because the person receiving it had died. The death adjustment factor for each type of income was calculated from 1970 census data as follows:

$$[\text{Death Rate (Y)}]_j = 1 - \left[\frac{\sum_i [.75 (Y)_{ij}] \times [MR_i] \times [\text{Adj. Recipients (Y)}_{ij}]}{\text{Agg. (Y)}_j} \right]$$

where:

$$[\text{Adj. Recipients (Y)}]_{ij} = \left[\frac{\text{No. Recipients (Y)}_{ij}}{[1 - (.75 (MR_i))]} \right]$$

and where:

i = the ith age-sex group

j = the jth governmental unit

Y = type of income

MR_i = U.S. mortality rate for the ith age-sex group

No. Recipients = Number of income recipients as reported in the 1970 census

Although it is recognized that mortality rates for age-sex groups vary among geographic units, such differences are not thought to be large enough to be important in an adjustment such as this. For qualifying places, zip code data will be used to make the Social Security and Veterans' payment estimates. Qualifying places are those which fulfill the following condition:

$$\left[\frac{\sum_{Z \text{ in } P} [\text{EXEMPT}]_{PZ}}{\sum_{Z \text{ in } P} [\text{EXEMPT}]_Z} \right] \geq .8$$

where:

Z in P = a zip code (Z) having at least one return filed from place (P)

EXEMPT_{PZ} = Number of exemptions in the intersection set of (Z) and (P)

EXEMPT_Z = Number of exemptions in zip (Z)

Thus, if at least 80 percent of people living in a city's zip delivery areas also live within the city's boundaries, then the city is a qualifying place. For such places we assume that the transfer amount per exemption for the city is equal to the amount for the city's zip delivery areas. This amount will be multiplied times the total number of exemptions of persons living within the city to derive the aggregate amount for the city.

Control totals for Social Security and Veterans' payments will be developed for the sum of the qualifying places in the county and for the nonqualifying parts. This will be done in accordance with the equations shown below:

$$\text{EQ.Z1} \quad [72 \text{ CEN(Y)}]_{QP} = \left[\frac{[A(Y)]_{QP}}{[A(Y)]_C} \right] \times [72 \text{ AMT (Y)}]_C$$

where:

$$\text{EQ.Z2} \quad [A(Y)]_{QP} = \sum_{Z \text{ in } QP} \left[\left(\frac{Z(Y)}{\text{EXEMPT.}} \right)_{Z \text{ in } QP} \times [\text{EXEMPT.}]_{QPZ} \right]_{QP}$$

$$\text{EQ.Z3} \quad [A(Y)]_C \text{ and: } = \sum_{Z \text{ in } C} \left[\left(\frac{Z(Y)}{\text{EXEMPT.}} \right)_{Z \text{ in } C} \times [\text{EXEMPT.}]_{CZ} \right]_C$$

$$EQ.Z4 \quad [72 \text{ CEN } (Y)]_{\Sigma QP} = \sum_{QP \text{ in } C} [72 \text{ CEN } (Y)]_{QP}$$

$$EQ.Z5 \quad [72 \text{ CEN } (Y)]_{\Sigma NQP} = [72 \text{ AMT } (Y)]_C - [72 \text{ CEN } (Y)]_{\Sigma QP}$$

where:

CEN (Y) =	Z (Y) =	AMT (Y) =
SS VP(E) VP(P)	SSA SS payments Vet. Admin. Ed. Pay. Other Vet. Admin. Pay.	CEN(SS) from EQ.C1 BEA VET. ED. Pay. BEA Other Vet. Pay.

and where:

QP = Qualifying place in County (C)

NQP = Nonqualifying part of County (C)

C = County (C)

Z in QP = a zip code (Z) having at least one return filed from qualified place (QP) in County (C)

Z in C = a zip code (Z) having at least one return filed from County (C)

EXEMPT_{QPZ} = Number of exemptions in the intersection set of (QP) and (Z).

EXEMPT_{CZ} = Number of exemptions in the intersection of (C) and (Z).

Once these control totals have been determined, then the Social Security estimates will be made as follows:

I. For Qualifying Places

$$EQ.S1 \quad [72 \text{ CEN } (SS)]_{QP} = \left[\frac{[72 \text{ CEN } A(SS)]_{QP}}{\sum_{QP \text{ in } C} [72 \text{ CEN } A(SS)]_{QP}} \right]_{QP} \times [72 \text{ CEN } (SS)]_{\Sigma QP}$$

where:

$$EQ.S2 \quad [72 \text{ CEN } A(SS)]_{QP} = \left[\frac{[A(SS)]_{QP}}{\sum_{QP \text{ in } C} [A(SS)]_{QP}} \right]_{QP} \times [72 \text{ CEN } (SS)]_{\Sigma QP} \times [\text{DEATH ADJ. } (SS)]_{QP}$$

$$EQ.S3 \quad [A(SS)]_{QP} = \sum_{Z \text{ in } QP} \left[\left(\frac{Z(SS)}{\text{EXEMPT}} \right)_{Z \text{ in } QP} \times [\text{EXEMPT}]_{QPZ} \right]_{QP}$$

II. For Parts of Qualifying Places

$$EQ.S4 \quad [72 \text{ CEN } (SS)]_{PQP} = \left[\frac{[72 \text{ CEN } A(SS)]_{PQP}}{\sum_{PNQP} [72 \text{ CEN } A(SS)]_{PQP}} \right]_{PQP} \times [72 \text{ CEN } (SS)]_{QP}$$

where:

$$\text{EQ.S5} \quad [72 \text{ CEN A(SS)}]_{\text{PQP}} = [72 \text{ A(SS)}]_{\text{PQP}} \times [\text{DEATH ADJ. (SS)}]_{\text{PQP}}$$

$$\text{EQ.S6} \quad [72 \text{ A(SS)}]_{\text{PQP}} = \left[\frac{\left(\frac{69 \text{ CEN(SS)}}{69 \text{ POP}} \right)_{\text{PQP}}}{\left(\frac{69 \text{ CEN (SS)}}{69 \text{ POP}} \right)_{\text{QP}}} \right] \times \left[\frac{72 \text{ CEN (SS)}}{72 \text{ POP}} \right]_{\text{QP}} \times [72 \text{ POP}]_{\text{PQP}}$$

III. For Parts Outside Qualifying Places

$$\text{EQ.S7} \quad [72 \text{ CEN (SS)}]_{\text{NQP}} = \left[\frac{[72 \text{ CEN A(SS)}]_{\text{NQP}}}{\sum_{\text{NQP}} [72 \text{ CEN A(SS)}]_{\text{NQP}}} \right] \times [72 \text{ CEN (SS)}]_{\sum \text{NQP}}$$

where:

$$\text{EQ.S8} \quad [72 \text{ CEN A(SS)}]_{\text{NQP}} = [72 \text{ A(SS)}]_{\text{NQP}} \times [\text{DEATH ADJ.}]_{\text{NQP}}$$

$$\text{EQ.S9} \quad [72 \text{ A(SS)}]_{\text{NQP}} = \left[\frac{\left(\frac{69 \text{ CEN (SS)}}{69 \text{ POP}} \right)_{\text{NQP}}}{\left(\frac{69 \text{ CEN (SS)}}{69 \text{ POP}} \right)_{\sum \text{NQP}}} \right] \times \left[\frac{72 \text{ CEN (SS)}}{72 \text{ POP}} \right]_{\sum \text{NQP}} \times [72 \text{ POP}]_{\text{NQP}}$$

where:

SS = Social Security payments

QP = Qualifying place in County (C)

C = County (C)

EXEMPT_{QPZ} = Number of exemptions in the intersection set of (QP) and (Z).

PQP = Part of qualifying place (i.e., part of a township located in the qualifying place.

NQP = A nonqualifying place.

Subcounty estimates of the remaining sources of income will be made by assuming the subcounty distribution for each income type is the same as the distribution of the indicated allocator. These amounts will then be added together and their sum raked to the county estimate for nonfarm self-employment plus "other income" as shown below:

$$\text{EQ.1} \quad [72 \text{ CEN (Y)}]_{\text{P}} = \left[\frac{[72 \text{ CEN A(Y)}]_{\text{P}}}{\sum_{\text{P}} [72 \text{ CEN A(Y)}]_{\text{P}}} \right] \times [72 \text{ CEN (Y)}]_{\text{P}}$$

$$\text{EQ.2} \quad [72 \text{ CEN A(Y)}]_{\text{P}} = \left[\frac{\sum_{\text{Y}} [72 \text{ U. CEN A(Y)}]_{\text{P}}}{\sum_{\text{P}} \left[\sum_{\text{Y}} [72 \text{ U. CEN A(Y)}]_{\text{P}} \right]} \right] \times [72 \text{ CEN (Y)}]_{\text{C}} \times [\text{DEATH ADJ (Y)}]_{\text{P}}$$

$$\text{EQ.3} \quad [72 \text{ U. CEN A(Y)}]_{\text{P}} = \left[\frac{[72 \text{ A(Y)}]_{\text{P}}}{[72 \text{ A(Y)}]_{\text{C}}} \right] \times [72 \text{ AMT. (Y)}]_{\text{C}}$$

$$\text{EQ.4} \quad [72 \text{ U. CEN A(Y)}]_P = \left[\frac{[69 \text{ A(Y)}]_P}{[69 \text{ A(Y)}]_C} \right] \times [72 \text{ AMT. (Y)}]_C$$

$$\text{EQ.5} \quad [72 \text{ U. CEN A(Y)}]_P = \left[\frac{[69 \text{ A(Y)}]_P}{[69 \text{ A(Y)}]_{QP}} \right] \times [72 \text{ AMT (Y)}]_{QP}$$

$$\text{EQ.6} \quad [72 \text{ U. CEN A(Y)}]_P = \left[\frac{[69 \text{ A(Y)}]_P}{[69 \text{ A(Y)}]_{\Sigma NQP}} \right] \times [72 \text{ AMY (Y)}]_{NQP}$$

$$\text{EQ.7} \quad [72 \text{ U. CEN A(Y)}]_P = \frac{72 \text{ AMY (Y)}}{72 \text{ POP}} \times [72 \text{ POP}]_P$$

where:

CEN (Y) =	U. CEN A(Y) =	Allocator A(Y) =	AMT. (Y) =	EQ. to EST. U. CEN A(Y) =
NFSE PLUS OTHER INCOME	INT.	72 IRS Interest	72 BEA Interest	3
	DIV.	72 IRS Dividends	72 BEA Dividends	3
	NFSE RENT & ROYALTIES ESTATE & TRUST PENSIONS	72 Estimates IRS C/E Income	NFSE & 72 BEA Net Rent & Royalties & Estates & Trust & Pensions	3
	VET. ED BEN			
	Parts of QP Other Parts of Co.	Veteran students receiving other income in 1969	[A(VP(E))] QP: from EQ.Z2 [72 CEN (VP(E))] ΣNQP : from EQ.Z5	5 6
	OTHER VET. PAY.			
	Parts of QP Other Parts of Co.	1969 Veteran Population	[A(VP(P))] QP : from EQ.Z2 [72 CEN (VP(P))] ΣNQP from EQ.Z5	5 6
	UNEMPL.	1969 No. Unempl.	72 BEA Unempl. Benefits	4
	FED. FELLOWSHIPS	1969 No. Grad. Stn.	72 BEA Federal Fellowships	4
	INDIAN TRANSFERS	1969 No. Indians	72 BEA Indian Transfers	4
	ALL OTHER TRS.		72 BEA Other Transfers	7

and where:

P = Subcounty part of County (C)

C = County (C)

QP = Qualifying place in County (C)

NPQ = Nonqualifying part of County (C)

APPENDIX

KEY

CODE FREQUENCY OF RETURNS

The "coding guide" technique for geocoding nonrespondents for income estimating purposes rests on two assumptions. First, that the geocodes on the fully coded IRS returns, are for the most part, correct. Second, that the geographic code frequency for a given address key is the same for nonrespondents as for respondents. The meaning of the second condition will be discussed shortly. If these two assumptions hold, then the procedure should yield estimates of the number of returns and income amounts which have the same expected value for a governmental unit as would have been the case if initially all returns had been fully coded.

A summary of the procedures is as follows: First, for all fully coded filers an address key is determined. This key consists of a 2-digit code indicating their State, their 5-digit Zip Code, up to 7 characters of the place name where their post office is located, and a code indicating their type of address. The following type of addresses were delineated: 1 = a numeric street address, 2 = a rural route, RFD, Star Route, etc. types of addresses, 3 = a post office box address, and 4 = all other types of addresses. Second, for each address key, all of the geographic codes associated with that key are arrayed along with their frequencies of occurrence. Third, nonrespondents are assigned one of the geographic codes associated with their address key in a random manner using the code frequencies as probabilities. For example, an address key may have four codes associated with it as follows:

2345196ABCDEFG1	231476293424	.73
	231476293426	.22
	23147660523	.03
	231491233723	.01

An uncoded return with this key would have a probability of .73 of being given the 1st code, a probability of .22 of being given the 2nd code, and so forth.

Let us now return for a moment to the second assumption of this procedure. It was assumed that the geographic code frequency for a given address key is the same for nonrespondents as for respondents. This assumption will hold, at least approximately, in all cases except when one code does not dominate a key and there is a sizeable variance in the response rates among the codes of the key. Thus, a dominant code will mitigate against the effect of different response rates for the codes in a particular key. To the extent that no code dominates a key, the assignment of geocodes to nonrespondent returns will be a function of differences in the response rates among the codes in that key. It is not possible at this time to assess the frequency or the magnitude of this problem and its effect on the estimates, however, several tables have been designed to measure the effects when the processing begins. Although this technique is still experimental, it is believed to be a promising approach to the vexing problems of assigning meaningful geographic codes to addresses.

Measurement of State and Local Tax Effort

John R. Coleman

Bureau of the Census

Three primary data elements are used in determining revenue sharing entitlements: population, income, and taxes. This paper will address itself to the composition of, development of, and use of the tax element and of a related element—intergovernmental transfers. To develop essential tax data, one must first answer the question, "What is a tax?". A rather simple and straightforward question it would seem on the surface. All of us pay taxes, and while some taxes have greater visibility than others, we suspect we can recognize a tax when we see one. Unfortunately the identification of tax amounts collected, and particularly the determination of which governmental unit has imposed the tax, is frequently a difficult matter.

Definition of Taxes

Let us first examine the definition of "taxes" as this concept is used in determining tax effort for interstate allocation of revenue sharing funds. Later, we will discuss the concept of "adjusted taxes" which is used for the intrastate distribution. Taxes are defined in Section 109 of the Fiscal Assistance to State and Local Governments Act (hereafter referred to as Revenue Sharing Act) as:

"... compulsory contributions exacted by the State (or by any unit of local government or other political subdivision of the State) for public purposes (other than employee and employer assessments and contributions to finance retirement and social insurance systems, and other than special assessments for capital outlay), as such contributions are determined by the Bureau of the Census for general statistical purposes."

"As such contributions are determined by the Bureau of the Census for general statistical purposes"—this is the phrase that restricts, that restrains, that draws the lines of limitation tightly around that most blasphemed of words—"taxes". To the anguish of many local government officials, the above qualification results in the exclusion of such compulsory contributions for public purposes as maintenance assessments, tax equivalents, sewer charges, garbage collection fees, and fines.

One may wonder why Congress adopted the Census Bureau's relatively narrow definition of taxes. The answer probably lies in large part in the immediate accessibility of the Census Bureau's highly respected public sector statistical series with its underlying national classification system during consideration of the Revenue Sharing legislation. The public finance series

extends back more than 70 years and provides a reasonably consistent application of standard definitions imposed on an extremely diverse universe of 78,218¹ local governments in addition to scores of thousands of dependent agencies, boards, commissions, districts, and authorities.

The Census Bureau's Governmental Finances publication, generally published in early fall of each year, presents State totals of State and local government tax revenues by level of government. These data are utilized by the Office of Revenue Sharing in formulae which calculate each State's entitlement.

State Tax Data

Two surveys are conducted each year to generate State government data. First, State tax aggregates² are developed by mailing questionnaires (Form F-5) to appropriate State tax officials shortly after the close of the fiscal year.³ The questionnaire requests a significant amount of detail for the various categories of taxes, i.e. property, income, general sales, specific sales, licenses, severance, inheritance, etc. As a reporting and editing aid, amounts for the previous year are included on the questionnaires. All returns are edited for consistency with aberrant entries receiving technical followup. The results of this survey are usually published in November in the State Tax Collections series.

One other tax element is required for the interstate distribution—State income tax collections. Since the Annual Tax Survey generates data on a fiscal year basis, and since by law,⁴ the collections used in the revenue sharing formula must be on a calendar year basis, required data are obtained by summing the results of four quarterly tax surveys conducted by the Census Bureau.

¹As of January 1972. Governmental Organization, Volume I, 1972 Census of Governments, U.S. Bureau of the Census, Washington, D.C., July 1973.

²State tax aggregates comprise amounts collected—including interest and penalties but excluding amounts refunded—from all taxes imposed by a State and collected by it or collected and transmitted to it by local governments. State tax revenue includes local shares of State-imposed taxes except any amounts locally collected and retained.

³June 30 for all States except 12-month period ending March 31 for New York, August 31 for Texas, and September 30 for Alabama.

⁴Public Law 92-512, Title I, Section 109(b)(2).

Local Tax Data

The local government tax aggregates are developed through an annual sampling of about 16,000 governments. Certainty survey coverage applies to all counties of 50,000 or more population in 1960, all cities of 25,000 or more population in 1960, all governments whose relative importance in their State, based on expenditure or debt was above a specified size, and a random sample of remaining units. Summation of data from certainty sample units generates aggregates which reflect over 75 percent of the amounts obtained in a census of the 78,000 governments. Computations of property tax revenue show a relative standard error of less than 2 percent for virtually all States and less than 1 percent for over two-thirds of the States.

Extensive efforts are made to keep nonsampling error at a minimum through substantial use of field and office compilation of large units and intensive edit of the smaller units. Questionnaires are edited for historical and internal consistency utilizing audit reports, various State reports, and many other secondary sources in the field of public finance to ensure the highest level of accuracy possible.

In regards to tax aggregates developed in this annual survey, it should be noted that taxes of all State and local governments (including independent special districts and school districts) are counted when measuring tax effort **only** for the interstate distribution.

Taxes for education purposes and taxes imposed by independent special districts are excluded when determining the intrastate distribution. Excluded taxes in Fiscal Year 1972 represented 25 billion of the 49 billion dollars⁵ imposed by local governments. Given the magnitude of excluded taxes, one can readily understand the dissatisfaction of officials of those governments whose citizens carry a heavy tax burden, much of which counts for naught in calculating intrastate entitlements because it is imposed for educational purposes or because it is imposed by an overlying or underlying independent special district.

Adjusted Taxes

Only general purpose governments, (approximately 38,500 counties, municipalities, and townships), are eligible for revenue sharing entitlements. To calculate intrastate entitlements a new concept was introduced, namely, adjusted taxes.

Adjusted taxes are taxes adjusted by excluding an amount equal to that portion of taxes which is properly allocable to expenses for education. Expenses for education are construed to include expenditures for current operation, capital outlay, and debt service. The boundaries of education are defined broadly enough to encompass primary and secondary schools as well as higher education,—vocational as well as academic studies. Before going further on the subject of adjusted taxes, it is important to realize that from the viewpoint of a local government, a larger adjusted tax is good, a smaller adjusted tax bad. Except where

the per capita limitations of the Act restrict entitlements, a larger adjusted tax generally produces a larger entitlement of revenue sharing funds. And given the competitive nature of the Act, what is good for one government within a county area is generally bad for one or more competing governments within that same area. Adjusted Taxes is a concept which presumes the ability to specifically identify school taxes; however, the realities of municipal finance question that presumption and present data handlers with some of their most difficult problems.

While there are 15,780 school districts with sufficient fiscal and administrative autonomy to be classified as separate independent governments, there are 1,457 school systems located primarily in New England and in the South which are classified as dependent upon a county, municipality, or town. For the most part the independent school districts have clearly identifiable taxes and thus present few data problems. Further, upwards of two-thirds of the 1,457 dependent school systems which have their school finances intermingled with general government finances, for one reason or another, maintain records which clearly identify school taxes. Massachusetts is an example of this situation where for a number of years State law has required such identification.⁶ However, the remaining one-third of the dependent school systems located primarily in Connecticut, Maryland, and Virginia generally cannot identify school taxes. For most units within these States an appropriation is made for school purposes out of a General or similarly named fund. The revenues to support expenditures of this fund come from a variety of sources including both tax and nontax revenue.

Original procedures for determining school taxes for such units consisted of identifying total school expenditures, identifying all specific school intergovernmental revenues, school service charges and other revenues with specific school identity, netting revenues against expenditures, and equating the residual amount with school taxes. This procedure came under attack in the early days of the revenue sharing program as opponents successfully argued that the procedure failed to adjust for cash surpluses used to finance schools or failed to permit a proportionate allocation of nontax revenues which carry no specific school identity.

Rectification of this problem came when the Office of Revenue Sharing issued a regulation (Section 51.21(b)(2)) revising the method of computing adjusted taxes. Called the RS-12 procedure after the form used in its calculation, it permits development of a ratio of taxes to nontax revenue with the ratio applied to school expenditures to produce a derived school tax amount. Beginning and ending of year cash balances are added or subtracted as appropriate from net revenues to allow for use of cash draw down or conversely, disallow use of cash build up. This procedure can only be used where school taxes carry no specific school identity.

In addition to school taxes imposed by independent school districts and by general purpose governments for dependent school systems, there are countywide school levies imposed in many States by counties (predominantly in the Midwestern and Western States) and paid intergovernmentally to the inde-

⁵ Governmental Finances in 1971-72, U.S. Bureau of the Census, Washington, D.C., November 1973, table 16, and special tabulation of adjusted taxes.

⁶ Annotated Laws of Massachusetts, Volume 2, Chapter 59, Section 23C.

pendent school districts. While these taxes are based on millage levied specifically for school purposes, care must be taken to ensure their exclusion from the county government's adjusted taxes.

School Debt Service

One further problem that exists in identification of school taxes relates to servicing school indebtedness. General purpose governments in some States—Alabama is a good example—issue debt in the name of the county or city and either construct schools or pay the bond proceeds over to independent school districts which construct their own schools. The debt service (payment of interest and reduction of debt) is generally financed from property taxes. Accordingly county or city taxes must be reduced by the amount of debt service to arrive at their adjusted taxes. This is a relatively easy operation when the debt carries an education identification. However, where school debt is collapsed into general purpose borrowing, e.g. for schools, fire stations, roads, etc. it is difficult if not impossible to determine how much taxes for debt service is attributable to education. With current record keeping practices the difficulty increases with each passing year's departure from the year of the original issue.

The problem will be compounded if interest rates in the municipal bond market reach a level where governments are motivated to refund outstanding debt to obtain a lower interest cost. Municipal finance officials would be remiss in their duties if at that time they did not refund school debt through general purpose bonds to remove school identity. It is likely that regulations will have to be written in the near future to remove the incentive to refund school debt into general debt.

Annual Surveys of Tax and Intergovernmental Revenues

During the past 18 months the Census Bureau has conducted three surveys to determine adjusted taxes and intergovernmental revenues covering three fiscal years (1971, 1972, and 1973) for the 38,500 general purpose governments. In the interest of minimizing respondent burden, the first two surveys made extensive use of the data obtained in the Annual Survey of Municipal Finances and in the 1972 Census of Governments. Supplemental questionnaires were dispatched to solicit data relating to school taxes as well as to complete the file for units that were not on the annual survey panel, or that had not responded to the earlier surveys.

The Fiscal Year 1973 survey used two basic questionnaires; a Form RS-9 dispatched to 34,000 governments, and a Form RS-12, dispatched to about 400 governments which finance schools through General Fund appropriations. Data for the remaining 4,000 general purpose governments were collected through cooperative arrangements with State officials in four States; Florida, New Jersey, North Carolina, and Wisconsin.

Fiscal Years

Compacting collection of data for 3 fiscal years into an 18-month period, while necessary to provide the most current data possible, increased the confusion in an already perplexing fiscal year situation. For each fiscal year, data are requested for the 12-month fiscal period that ended between July 1 and June 30. Of the 3,044 counties in the United States, 52 percent have December fiscal years, 35 percent June, 7 percent September and 6 percent are spread over the remaining months. Fiscal years of municipalities and townships have even greater dispersion. Add to this mixture, dependent agencies which have different fiscal years than their parent government; sprinkle in four entitlement periods, three covering 6 months, one covering 12 months; add several planned use and actual use reports covering differing periods; season with thousands of part-time nominally paid local government officials, and the result can reasonably be expected.

Confused fiscal year reporting problems to date have been second only to school tax identification problems. There is, however, hope for a diminution of the fiscal year problem in the future. Intensive efforts to instruct local officials regarding correct fiscal periods; a spreading of data requests to once in a 12-month period; and synchronization of Census Bureau/Office of Revenue Sharing contact with local officials—all can be expected in the future to lessen considerably misinterpretation of fiscal periods covered.

Agency and Shared Taxes

A problem area demanding significant editing attention is the reporting of shared taxes or agency taxes as one's own tax effort. A local government can only receive credit for taxes which it imposes itself, with the single exception that State imposed, locally collected and retained taxes are credited to the local government. Other than this exception, it is irrelevant which government collects the tax. Frequently a county collects property taxes for all overlying and underlying governments; in such an instance, the county can be credited only for those local taxes imposed by its legislative body. For other taxes, a county is viewed simply as acting in an agency capacity. At a higher level a State government may collect a county imposed sales or income tax which is piggybacked on a State sales or income tax and remit the proceeds to the county. In such a case the State is acting in an agency capacity for the county imposed portion.

For the hundreds of State imposed taxes which are shared locally,⁷ the local governments must report such amounts as intergovernmental revenues—not as taxes. In collecting data on taxes and intergovernmental revenue, experience has proven

⁷ See *State Payments to Local Governments*, Volume 6, 1967 Census of Governments, U.S. Bureau of the Census, Washington, D.C., December 1968, for a detailed analysis of shared taxes. The corresponding publication from the 1972 Census of Governments is scheduled for publication in February 1974.

that very detailed special instructions must be provided local officials specifying whether a tax reflected in his records is: (1) a tax imposed by his government and therefore reportable in adjusted taxes, (2) a tax for which his government acts in an agency capacity and therefore nonreportable, or (3) a shared tax and therefore reportable as an intergovernmental revenue.

Memphis Rule

One exception exists to the Census Bureau's procedure of attributing local taxes to the imposing government. Section 109(e) of the Revenue Sharing Act (known as the Memphis Rule) provides that any county imposed sales tax exacted from a municipal or township area and shared with the municipal or township governments without specifying purposes for which the unit may spend the revenues shall be treated as taxes of the receiving government. The Governor must first certify that such a condition exists before the rule can be invoked.

Memphis Rule adjustments have been made to date in seven States: Alabama, Louisiana, Nevada, New York, North Carolina, Tennessee, and Virginia, with the largest adjustments encountered in Tennessee and in the upstate counties of New York. The Memphis Rule is so named because of its applicability in Memphis-Shelby County, Tennessee, where an adjustment of approximately 12 million dollars in adjusted taxes was required in Fiscal Year 1971.

Note that the Memphis Rule exception applies only to county imposed sales taxes, not State sales taxes, not city sales taxes, not county income taxes, not county property taxes. The rationale for singling out sales taxes for special treatment—county imposed sales taxes at that, is difficult to follow if promoting greater equity was a motivating factor in its adoption. It may be conjectured that there was not widespread Congressional awareness during consideration of the bill of the extent to which governments share taxes which they impose, but which they extract from inhabitants of the shared government's area. This may be an issue that receives greater attention when the revenue sharing program is subjected to Congressional analysis for extension, modification, or termination in the future.

Intergovernmental Revenues

A rather innocuous little paragraph in Section 108 of the Act imposes one of the most burdensome data collection responsibilities of the program. Subsection (b), paragraph 6(C) reads in part:

"Limitation—The amount allocated to any unit of local government under this section for any entitlement shall not exceed 50 percent of the sum of (i) such government's adjusted taxes and (ii) the intergovernmental transfers of revenue to such government...."

In effect this paragraph requires that every intergovernmental dollar passing between and among local governments be identified. This is a monumental task. It first requires that all agencies, boards, commissions, districts, authorities, and departments be properly classified as either governmental or non-governmental and for those which are governmental, that they be classified as dependent upon a specific government.

Local government officials generally tend to carry a narrow or restricted view of the scope of their government's activities. The hospital authority, joint health department, city-county airport authority, etc. are frequently viewed locally as not being part of one's government. In actuality these agencies may meet all necessary criteria for classification as dependent agencies of governments which deny or are unaware of their dependency. Extensive use is made of supplementary instructions in Census data collection programs to inform local officials of the full scope of their government's activities.

A multitude of quasi-governmental agencies were created during the 1960's, often as a result of the profusion of categorical grant programs which came into existence; Community Action Agencies, Drug Rehabilitation Organizations, Planning Agencies, and Councils of Governments are a few of the type of agencies through which significant governmental activities are carried out, but which may or may not be governmental units. Failure to identify a Community Action Agency (CAA) which meets criteria for classification as a county agency would result in understatement of the county's intergovernmental revenues by the amount of Federal and State aid which the CAA receives.

The need for complete identification of intergovernmental revenues is particularly acute in Southern counties, as a number of these units are subject to the 50 percent limitation; i.e., 50 percent of the sum of their taxes and intergovernmental transfers exceed the amount for which they are entitled; hence the excess is lost by the county government and added to the State government's entitlement. It becomes obvious that for those governments up against the 50 percent ceiling, if a \$100,000 Federal welfare grant is inadvertently unreported, the county could lose up to \$50,000 in its revenue sharing entitlement. The true potency of data in the revenue sharing program manifests itself when reporting errors of this nature are analyzed.

Comparing Noncomparables

If one were to attempt to describe in one word the most striking feature of revenue sharing, the word would have to be "competitiveness". Quite probably there are few programs which have ever caused officials of local governments to compare themselves with other governments to the degree that revenue sharing has. Surface comparisons made by competing governments have often focused on population, an intuitive sense of relative income levels, and total property tax collections or millage imposed. These comparisons frequently fail to consider the tremendous diversity that exists both in organizational structure of the governments and in the financing vehicles utilized.

For example, one government may make greater use than another government of public housing which generates in lieu of tax payments rather than property tax payments. By definition, "in lieu of taxes" do not qualify for inclusion in adjusted taxes. Another government may finance garbage collection from taxes; another government, from service charges, or that function may even be performed by private firms. Service charges are excluded from adjusted taxes. Another government may use maintenance assessments or special assessments for a given purpose; the next may use property taxes—the assessments can not be counted. One city may have its Public Works Department

perform a function financed primarily from taxes. In a neighboring city the function may be performed by the county or by an independent special district. The latter city fails to get credit for the tax effort that finances this function.

And another government may operate a water utility, charging higher rates to produce a profit which is then used to finance general government. The profit, while having the affect of tax, cannot be counted. However, a neighboring government may operate its utility on a break-even basis, but levy taxes at a rate designed to produce revenue at the same level as the profit producing utility. The latter would be able to count its taxes and thus obtain a higher adjusted tax. The list of "inequities" is endless, and the organizational and financing diversity is even greater when making interstate comparisons.

Tax Relief

An area with the potential for introducing major inequities into the revenue sharing program is that of the property tax relief programs which have exhibited astronomical growth in the past several years. Hundreds of millions of dollars are involved in these tax relief programs. Three basic methods are in use today to implement tax relief. First, local governments reduce property taxes of their citizens and thus forego collection of taxes. States reimburse the local governments for the taxes foregone. This procedure produces no double count in the State-local tax effort.

Secondly, local governments impose and collect the taxes from their citizens. State governments then permit the taxpayers to reduce their State income tax liability by the amount of the tax relief. This procedure also produces no distortion because State income taxes are reported net of the tax relief refunds.

Thirdly, the situation is the same as the second except the income tax is not used as a vehicle for granting relief. Instead, the State makes a direct expenditure rather than a tax

expenditure. It sends a check to each taxpayer eligible for tax relief. This method does distort in that it counts both the full local taxes and the full State taxes. There is no reduction for tax relief. States which opt to use this last method gain a competitive advantage over other States insofar as tax effort determines the proportion of Federal funds to which each State is entitled. This last procedure is politically useful to executive officers as they can receive political credit from local taxpayers for the full amount of the tax relief check. This has greater visibility than the netting procedure used in conjunction with the income tax refund procedure. In view of the size of the mushrooming tax relief programs, it is probably reasonable to expect that special regulations will have to be developed to deal with this problem.

Conclusion

Thus, it can be seen that frequently the "maldistribution of entitlements" (at least as perceived by local government officials who feel "wronged") has its roots in organizational relationships created and fiscal mechanisms utilized in pre-revenue sharing days—some created and utilized as a result of extensive deliberation; others decided almost by chance, by the political vicissitudes that at times shape a government's course of action.

With the growing awareness of the critically important role that data play in the revenue sharing program, a number of local governments, operating within constitutional, statutory, political, and economic restraints, have begun to weigh trade-offs of various organizational realignments and modification of financing vehicles in order to maximize the one data element that is most within their capacity to affect; i.e., adjusted taxes.

It is still too early to assess the full extent of changes that will occur as a result of the data intensive nature of the revenue sharing program, but one can safely say that the statisticians will be monitoring changes closely to ensure availability of a data file that is both current and accurate. Given the magnitude of the funds being distributed, nothing less will suffice.

Statistics of Revenue Sharing After 1 Year

Arthur L. Hauser, Office of Revenue Sharing

I'm pleased to have the opportunity to talk to you about our impressions of data-related changes that have occurred in the period slightly over 1 year since general revenue sharing was born.

After the passage of the State and Local Fiscal Assistance Act of 1972, better known as the Revenue Sharing Act, in October 1972, Robert P. Strauss, then consultant to the Secretary of the Treasury, wrote: "The historic action of Congress to create general revenue sharing represents not only a revolution in the form of American federalism, but also a revolution in the Federal statistical base that will be used to administer the program." (Nation's Cities—October 1972)

In my remarks today, I would first, like to consider the so-called "revolution to the Federal statistical base" that Mr. Strauss expected; the kinds of changes to State and local governments' data that we can see happening through our day-to-day work with the Bureau of the Census. In addition, I'd also like to touch on general revenue sharing's impact on the Bureau of the Census, as I see it. It is apparent from our daily contacts with Census personnel that the data needs of revenue sharing have exposed them to frequent inquiries from many small units of local government—units that previously were not in their survey of finances or units that had reported previously only in the Census of Governments every 5 years.

Revenue Sharing's Influence on the Data

The influences of general revenue sharing on the State and local government's data can be categorized in two ways. The first relates to activities which are likely to improve or change that data from sources which had been in existence prior to the advent of revenue sharing. The second relates to the availability of new data sources, those not existing prior to revenue sharing.

DATA IMPROVEMENT

Prior to revenue sharing, the Census Bureau's activities with local governments consisted of the previously-mentioned survey of finances to about 16,000 of the 78,000 governmental units. As a voluntary periodic survey, it was subject to the problems familiar to most of you—involving mostly response and statistical significance.

As a result of general revenue sharing, there are many specific mechanisms which are operating to bring about improvement in the Census-produced data used by the Office of Revenue Sharing. This is especially true of the revenue data—adjusted taxes and intergovernmental transfers. In the case of the revenue data, the improvement process begins at the doorstep of the State and local government officials who either complete survey forms or certify to their accuracy. By now, most officials of recipient governments generally understand that the information that they report on the Census Bureau revenue sharing survey forms has a direct bearing on the amounts of revenue sharing funds that they receive. This understanding has contributed to revenue data in the following ways:

1. The response rate to Census Bureau surveys of revenue data for revenue sharing has improved considerably since the Special Revenue Sharing Survey to obtain Fiscal Year 1971 revenue data. At the time of our first allocation, the Special Revenue Sharing Survey to obtain this data had a response rate of approximately 75 percent. Strenuous efforts to increase this response rate after the allocation were made by Census Bureau. Many governments who had not responded initially did so once the revenue sharing payments had begun. This resulted in a final response rate of about 93 percent; it also meant that it resulted in more than 5,000 data changes from the estimated data which were initially used by our office. In the 1972 Census of Governments which was supplemented by some special revenue sharing forms, the response rate for Fiscal Year 1972 revenue data rose to approximately 98 percent. Presently, the Special Revenue Sharing Survey of Fiscal Year 1973 revenues is preceeding according to plan. Thus far, a response rate of approximately 97 percent has been achieved.

2. Local and State government officials responsible for completing the survey forms have increasingly sought guidance from the Office of Revenue Sharing and the Bureau of the Census. The Bureau of the Census, in particular, has worked hard and spent considerable time responding to these individual requests for help and, in addition, has provided more detailed special instructions than heretofore covering the unique revenue situations of different State and local governments.

3. The survey forms sent to State and local officials by the Bureau of the Census are now more frequently reviewed by local officials with a higher level of responsibility than was previously the case. Prior to this, the survey form was in many cases returned directly by a town clerk or other clerical personnel without review by higher officials.

Once the raw revenue data are received by the Bureau of the Census, the Governments Division makes careful editing checks to insure its accuracy. Because the Office of Revenue Sharing has on several occasions become aware of some data differences which were large enough to have an impact on the allocations, we have worked closely with the Bureau of the Census to positively identify the existence of an error or to verify the data. The increased use of and refinements in editing checks have already given us more accurate revenue data.

When the Census Bureau completes its processing of the data, they provide us with the new set of data. This new data is used by us to calculate recipient governments' allocations. However, the efforts to improve the data do not stop at this point. The Bureau of the Census has an ongoing program to review the quality of submitted data. This review occurs as a result of the submission of revised reports from local governments and additional material received from State and other sources. Furthermore, the Office of Revenue Sharing's Regulations specify that the data used in the allocation must be made available to all units of government to give them an opportunity to challenge the data for their government if they believe it to be in error. Thus, through our administrative procedure now known as the data improvement program, many governments have pointed to data differences—and have supplied documentation for review by the Census Bureau.

Our first data improvement program was in December 1972 when we sent the 38,000 recipient governments the data used to calculate their revenue sharing allocations for the first entitlement period. Approximately 4,000 governments raised questions about their data in the following months although fewer than half of these appeals were found to have sufficient merit to warrant a data change. This process did much to bring about a better understanding of the data elements (population, per capita income, adjusted taxes and intergovernmental transfers) on the part of all recipient governments. We have just concluded our data improvement program covering the data used in the 4th entitlement period. This time, only 2,000 governments challenged; approximately 1,100 challenges required research while the balance were submitted without documentation. Of the 1,100 inquiries, about 82 percent resulted in revenue revisions.

The government address information generated by the Office of Revenue Sharing also contributed to the data improvement. Many governments provided us with address corrections. The Bureau of the Census also has used some of this new address information to improve its address records which are used in sending out survey forms. Obviously, there is a better chance of obtaining the desired data if the survey forms are sent to the proper official for completion.

SOME DATA PROBLEMS: SOURCES OF FUTURE DATA IMPROVEMENT

Although there has been significant improvement in the data used for revenue sharing, there are reasons for believing that the data are not as good as we would like them to be and, therefore, that there is room for continued improvement. For example, it is likely that some governments may not have reported data discrepancies when it was not to their advantage to do so. Thus,

adjusted taxes carried on the high side or per capita income carried on the low side are more likely to go undetected than adjusted taxes or per capita income in the other direction. Also, a few governments may be tempted to exaggerate the amount of taxes that they report, rather than the amount actually collected. We expect, though, that the various checks employed by the Census Bureau should keep this situation under control.

A few adjusted taxes reported on the high side have been discovered due to challenges of a governments' data by other governments which are competing with them for revenue sharing funds. We have not encouraged this type of challenge because we have not wanted to emphasize unduly the competitive aspect of the formula. However, the irony of the situation is that our experience shows that this type of challenge is not likely to be advantageous to the particular unit of local government raising such a challenge. This is because the reduction of a competing government's adjusted taxes data will result in a reduction to the county area's adjusted taxes. This, in turn, will result in a reduced allocation to the county area.

We will, of course, work with the Bureau of the Census to correct all data differences regardless of who discovers them, because this will contribute to greater equity in distributing revenue sharing funds. We expect that our ability to detect and make necessary changes will improve with time.

NEW SOURCES OF DATA

Besides improving existing data sources, general revenue sharing has required development of new sources of data. Let's look at a few of these.

1. **Revenue Data**—A Special Revenue Sharing Survey of all general purpose governments will be completed annually except when the Census of Governments is taken to compile data on adjusted taxes and intergovernmental transfers for revenue sharing. The use of the adjusted taxes (taxes minus educational taxes) data in the revenue sharing formula has made it necessary for the Bureau of the Census to develop a new survey form to obtain the data in the desired manner for the program. Perhaps, of more significance is the establishment of communications with all general purpose governments—the bulk of which have populations under 2,500. Of approximately 38,000 general purpose governments, approximately 27,000 have populations under 2,500. And approximately 35 percent of the 38,000 governments (13,500) have populations under 500. Thus, general revenue sharing has brought about a new dimension in State and local government revenue data.

2. **Population and Per Capita Income Data**—For general revenue sharing, both the 1970 population and per capita income data are adjusted annually to reflect boundary changes for places of 2,500 or more based upon information received by the Bureau of the Census in the annual Boundary and Annexation Survey. Before revenue sharing, information from the Boundary and Annexations Survey had been used by the Census Bureau only to keep their Census maps up-to-date.

Still on the subject of population and per capita income, you have heard how Census Bureau has the challenging job of updating these data for local governments using administrative records.

3. **Population for Indian Tribal and Alaskan Native Village Governments**—Although the Bureau of the Census has collected data on Indians and Alaskan native persons in the United States, it has with few exceptions never collected data specifically on Indian and Alaskan native population which are members of and under the jurisdiction of Indian tribal and Alaskan native governments. There has never been a need for this data before. Because this type of population data was needed for revenue sharing, estimates of these populations have been made by the Bureau of Indian Affairs for the Office of Revenue Sharing. Currently, we are exploring other avenues for obtaining updated population data for Indian tribal governments and Alaskan native villages.

IMPACT ON THE BUREAU OF THE CENSUS

I've already noted some of the ways general revenue sharing has brought important changes in data. And we know that this means that the Census Bureau has successfully made some changes.

We know that general revenue sharing is not entirely unique to the Bureau of the Census because Census data is used by other Federal agencies in conjunction with the granting of funds to State and local governments. Yet the Census data used for general revenue sharing is of special importance because the data directly determines the amounts that recipient governments will be allocated.

Because the data have this kind of special importance, especially the revenue data, and because the Office of Revenue Sharing must insure the equity and correctness of our allocations in **all** cases—not merely on the average—we necessarily view the data not as statistics but as ledger entries in each government's account. We view each revenue sharing recipient as a separate "account" as differentiated from all other "accounts." Thus, our criteria for the data are different and our demands for data are those of the accountant.

Given these demanding criteria, we see the Census Bureau demanding—and getting—more precise revenue data from respondents. This is happening because there is a recognition that the revenue data are used in a direct, programmatic formula to determine entitlements. Further, such recognition means that it is necessary to achieve high-quality data down to the smallest area level. But we are not excluding the value of estimates and sample data. As you will note by reading our other material here, we use the Census State population estimates and tax effort estimates for the interstate allocation.

In conclusion, we feel that those persons like Bob Strauss who anticipated that administering the general revenue sharing program would bring about very substantial changes in the kinds and quality of statistical data were correct.

I would also like to point out that we recently turned over to the Census Bureau, who will be acting as our agent, magnetic tapes containing the final data for entitlement periods 1, 2, and 3. For those of you interested, please contact the Census Bureau's Data User's Service Office regarding sale.

Making the New Federalism Work: Goals for General Revenue Sharing's Data Needs

Robert P. Strauss*
Associate Professor of Economics
University of North Carolina at Chapel Hill

I. Introduction

I want to express my great pleasure at being with you today. The passage of general revenue sharing portends a new era in micro-data needs. I should note that the general movement towards revenue sharing and away from categorical or conditional grants-in-aid, otherwise known as the New Federalism, will require a variety of micro-data sets which are frequently updated.

General revenue sharing has highlighted the need for recurrent measures of population, community money income, and local tax collections. Not only is the "recurrency" worth noting but also the logical unit of analysis, namely general local governments. If we look down the road towards the passage of special revenue sharing, we should be able to anticipate that additional micro-variables by the same as well as other (special) geographic subdivisions will be necessary. Of course, we do have some of this data, especially for more populous areas. But my point here is that the data requirement for smaller areas have yet to be met.

A corollary to this is that with the advent of the New Federalism, more rather than less will (or should) be spent by the Federal Government for micro-statistics. However, I want to indicate that not only should more be spent for collection and upgrading of new and current series, but serious attention be increasingly given to verify such self-reporting data to ensure accuracy. There are many areas in which formula design will depend on data or behavior which ultimately is at the control of a locality or group of recipients. The use of adjusted taxes in general revenue sharing is just one example. There are many other areas in which we might want to reward or dissuade organizational behavior for certain outcomes; however, sufficient verification (as well as careful definition) must occur to ensure that program goals are met.

Having preambled my remarks with the prognostication that the New Federalism will inevitably require more micro-data and accordingly require greater Federal resources to meet properly these increased needs, let me turn to the data implications per se of general revenue sharing. Last year I noted that a revolution in Federal statistical base was implied by the State and Local Fiscal Assistance Act of 1972; my theme today is that successful revolutions require forthright leadership, for the

successful changes in the revenue sharing data base will directly affect the success of the program. I shall discuss in turn the leadership issues relevant to income, population, adjusted taxes, and the formula.

II. The Data Issues of General Revenue Sharing

A. Income Updates

During Executive Session deliberations of the House Committee on Ways and Means, repeated discussions took place on the general problem of updating the data elements ultimately used in the legislation. As you may recall, an overhaul of the welfare system was being contemplated which, coupled with changes in reporting requirements within the tax system, might yield post-censal data on total money income by jurisdiction. I do not think it is a breach of confidentiality of those proceedings to reveal now that the Committee seriously considered requiring that the welfare system contain county and incorporated place of residence as it ultimately did of those in the tax system. Because the welfare reform was pending, and because it was impractical to amend the existing Social Security Act, Undersecretary Veneman testified that costs of several hundred million would be involved as well as considerable inconvenience to modify existing reporting systems. Consequently, no legislative action was taken to gather income data about low income persons (or nonfilers). However, it was clear from Veneman's testimony and Committee sentiment, that eventually this data ought to be collected as it was their clear determination that such residence information be collected of taxpayers to facilitate estimation of money income by political subdivision.

Of course, it is now history that the 1972 1040 form required such information and is to be used to obtain income updates. Several things, however, ought to be underlined about this bit of legislative history. First, the tax forms were to be permanently modified. Second, the welfare system as it is modified should be used to provide data about low income persons.

*Formerly Special Assistant to Deputy Treasury Secretary of the U.S. Treasury, Charles E. Walker, and recipient of Treasury Department's Exceptional Service Award for work on General Revenue Sharing.

More recent history indicates that the commitment to change the 1040 form and to require the welfare system to report has weakened and may vanish. Apparently the Commissioner of the Internal Revenue Service (IRS) does not interpret Section 144 of the State and Local Assistance Act to mean, as I take it to mean, that incorporated place of residence must be recorded and put on tape by the IRS each year. To those of us committed to making the New Federalism work, this can only be viewed as a setback. Also, the Federal assumption of the adult welfare categories has occurred without the requisite modifications of the reporting system. While the default this year (CY73) in obtaining incorporated place of residence may be viewed as a temporary setback to the statistical community, I should note that likely tax reform in 1974 that involves tax simplification, as it undoubtedly will, may well involve the complete removal of the residence questions from the tax form. Moreover, if income maintenance replaces the existing program for Aid to Families with Dependent Children, it is likely that place of residence will not be obtained unless substantial prodding takes place.

Without going into the debate on the nature of the question that was asked on the 1972 form (let me parenthetically note here the opinion that a 70 percent response rate to the question is rather high in view of the initial 40 odd percent response rate to the zip code questions that were asked in the mid sixties with greater attending IRS publicity), I would like to comment on the definitional problems that occur when comparing adjusted gross income to total money income. During those Executive Session deliberations Congressman Vanik suggested that total money income be reported on the 1040 tax form. It is clear that a Hague-Simons concept of income is what a tax-writing committee ought to use as a reference point when analyzing changes in the IRS code, not because that is the tax base, but rather because it is a sound point of comparison for equity purposes. While sentiment apparently existed for more complete reporting of nontaxable or partially taxable income, the thought was then that while desirable, it might be better effected through separate legislation. My point here is that we may well see a broader income concept reported through the tax system which would reduce the definitional problems rather considerably in obtaining total income of particular places.

There is, then, in the area of income data update, room for improved statistical leadership that will ensure that the original vision of updated and improved local income data will occur. This is not to discount the rather substantial difficulties of convincing the Internal Revenue Service and Department of Health, Education, and Welfare that improved local income data is worthwhile. However, it is significant that the legislative history of general revenue sharing contains the elements to make continued progress in this area.

B. Population

The second area of data updates involves population. Here we are better acquainted methodologically with the problems and prospects of post-censal updates than, I think, in the micro-income area. I would like to confine my remarks here to the area of undercounts that occurred and the extent to which updates in the form of corrections should be made.

Again, reference should be made to the legislative history and the ingredients in the legislation to obtain guidance on how the problem should be solved. During Executive Session deliberations that were ultimately published (April 17, 1972), then Director of the Census Bureau, George Brown and Congressman, William Green of Philadelphia, discussed various statistical problems, among which was the undercount of certain groups in core cities. Again, the sentiment of the Committee was fairly clear: at that time, while the extent of the undercount was still being researched but was as yet unknown, the general feeling was that when adequate figures were available, corrections ought to be made.

We now know the national undercount rates by age, sex, and race, and it strikes me as appropriate to use this data to adjust the population figures. I have done this for several States, and while better data will undoubtedly be generated (State by State rates, for example) it would seem appropriate at this point to account for what we know to have occurred, and then, at a later date, improve our statistics further when possible. I should note that the Secretary of the Treasury has authority to use such estimates, and such updates would be done with attending modification of regulations to provide that retroactive adjustments in payment not be made. Moreover, there is precedent for corrections as corrections have already been made to the census in the income area.

Again, the legislative history is clear: errors were anticipated and the Secretary empowered to use such updated or corrected data to provide corrections. And again, statistical leadership should do this expeditiously. I should note that the National League of Cities-Conference of Mayors has twice recommended that the undercount be corrected.

C. Taxes

We have more experience now with the adjusted tax data updates than with income and population. It is my judgment that the ability to update the tax data annually suggests the other two elements can be updated and corrected as well.

The points I want to stress with the tax data involve independent verification of the numbers and a hard and fast attitude concerning the definitions.

As is obvious, updating just the tax data has rewarded in entitlement terms communities with increased taxes which are below the 145 percent limitation. The nature of the data collection and verification process is clearly vital to a fair administration of the program. The data improvement program of the Office of Revenue Sharing (ORS) is a step in the right direction to ensure that the tax data is accurate; however, independent checks need to be made on local determination of taxes as adjusted by the Governments Division (Census Bureau) and then ORS to ensure that no inflation of figures is taking place. The incentive to exaggerate is rather obvious as is the incentive to report gross instead of net (of refunds) collections. State level verification is not uniform nor do all localities have local CPA certification. Leadership in this area could well involve encouraging the States to do audits and/or encouraging local CPA certification. For the budding revenue sharing research industry, some independent verification of taxes might prove useful.

The second issue involves the definitions themselves. The use of taxes rather than revenues has been criticized by some and suggestions for broadening the definition or moving to revenues instead have been made. Several points are in order: First, the choice was a deliberate one made by the Congress; and second, it resulted from an exhaustive interstate and intrastate analysis of the effects of the two concepts. Finally, while it may affect rather dramatically certain municipalities financed by utility operations, I think we have to examine rather carefully the geo-equity implications of such a revenue structure as well as the more esoteric, but nonetheless important issue of whether or not that is a proper role for the local public sector to play. Without discussing these in any greater detail, let me indicate my own personal reservations on both these issues.

D. Availability

General revenue sharing has been with us for more than a year. It was planned to be a grant-in-aid program characterized by a small central administrative office and a maximum amount of openness about the manner in which it is administered. I think the first goal has been reached, but worry about the second. Computer tapes of the basic data are still not available to the research community, and there is no prospect of their being available in the foreseeable future. While printed copies of the data are available, and I should note parenthetically that this is a step in the right direction, the rather crucial interstate data has yet to be published. Most would seem to be available from various Census Bureau and Bureau of Economic Analysis sources; however, it strikes me as entirely reasonable that the interstate data be published as well as the intrastate data. There is, then, room for additional leadership to make the data more available.

III. Formula Issues of General Revenue Sharing

My remarks so far concern data-related issues of general revenue sharing. A few points about the formula are in order as well.

A. Formula Availability

While the data is available in hard copy form, researchers as well as recipient units remain mystified about how the formula actually, that is, computationally works. The program(s) that create the intrastate allocations have not been made public to date, and I think some leadership in this area would be desirable. The incentives in the legislation require local awareness which in turn requires they be informed of how it works.

Making the program tape available was always contemplated by the Congress. It strikes me that time has now come to do this.

B. Floor Considerations

As I noted in San Juan at the December, 1973 National League of Cities Meetings, it is my understanding that the application of the 20 percent floor differs somewhat from the original algorithm; some quick research done indicated a \$6 million annual difference for Chicago. This may explain the rather dramatic change in allocations to townships in the Midwest. I suggest to the interested a comparison for rural Midwest counties of percentage of funds going to townships as reported by ORS and those reported in, say, the Senate Finance Committee Report.

IV. Conclusions

From my perspective, general revenue sharing has been and will continue to be a huge success. The funds are being spent for locally identified priorities with increased citizen participation. The incidence of fraudulent use of funds has been miniscule, and I conclude, far less than in those categories of grant-in-aid which were "outside" the local budgetary process.

The statistical burden of general revenue sharing has been enormous, but one that has been well-shouldered to date. Creative leadership in the future is absolutely essential for continued success of the program and the subsequent passage of the special revenue sharing bills. It strikes me as crucial that the Federal Government work closely with the States to encourage them to better monitor local micro-statistical efforts. State interests are vital in this, but they need to be led if not pressed by the Federal Government. Correlative to this is a more significant commitment by the Federal Government to maintain the spirit of the New Federalism. Public availability of data and formula are part of this. Continued progress in the use of existing Federal information devices, such as the determined use of individual income tax for population estimation purposes, is another. Finally, we must anticipate our micro-data needs of the future and press for additional micro-data sources. As noted before, planned revamping of the welfare system can provide us with income and population data on a geodisaggregated basis if the Federal Government can be convinced of their utility. The rationalization of our Federal system of government, indeed our general capability of social problem solving, demands a fore-sighted and co-ordinated micro-data program. It might be quite useful for the American Statistical Association in concert with the Office of Management and Budget's Statistical Policy Division, the Census Bureau, and the Federal Statistics User's Conference to organize a continuing discussion of micro-data needs for the coming grant-in-aid consolidations.

Formulas For Allocation of Funds Under the General Revenue Sharing Act of 1972 (PL92-512)

prepared by
Albert Mindlin, Chairman,
Committee on Small-Area Statistics
American Statistical Association

General revenue sharing funds are allocated to the numerous units of government in the following sequence.

1. The first allocation is made to States. Two formulas are employed. The formula yielding the largest amount for a State is usually used, subject to the overall limitation on funds.
2. The second allocation, a bookkeeping operation, is to an artificial geographic entity called a "county area."
3. The third allocation is to Indian tribes that exist in the "county area."
4. The fourth allocation is to county government within the "county area."
5. The fifth allocation is a bookkeeping operation setting aside aggregate funds for all townships that exist in the "county area."
6. The sixth allocation is to all units of local governments in the "county area" other than counties and townships.
7. The seventh allocation is to the townships.

This report exhibits the formulas for each of these allocations. Although the formulas can be algebraically simplified, they are exhibited in the same form in which they are described in the Act.

State Allocation

- P_i = population of i^{th} State
 $\sum_i P_i$ = U.S. population
 U_i = urbanized population of i^{th} State
 $\sum_i U_i$ = U.S. urbanized population
 M_i = aggregate money income of State (census definition of income)
 $\sum_i M_i$ = aggregate U.S. money income
 T_i = aggregate net State and local taxes of i^{th} State
 $\sum_i T_i$ = aggregate U.S. net State and local taxes
 R_i = aggregate personal income of i^{th} State (BEA definition of income)
 $\sum_i R_i$ = aggregate U.S. personal income
 K_i = aggregate net individual income tax collection of i^{th} State
 $\sum_i K_i$ = aggregate U.S. individual income tax collections

3-factor formula

$$\frac{P_i \cdot \frac{T_i}{R_i} \cdot \frac{\sum_i M_i / \sum_i P_i}{M_i / P_i}}{\sum_i P_i \cdot \frac{T_i}{R_i} \cdot \frac{\sum_i M_i / \sum_i P_i}{M_i / P_i}}$$

This can be simplified to:

$$\frac{P_i \cdot \frac{T_i / R_i}{M_i / P_i}}{\sum_i P_i \cdot \frac{T_i / R_i}{M_i / P_i}}$$

5-factor formula

$$\frac{P_i}{\sum_i P_i} + \frac{U_i}{\sum_i U_i} + \frac{P_i \cdot \frac{\sum_i M_i / \sum_i P_i}{M_i / P_i}}{\sum_i P_i \cdot \frac{\sum_i M_i / \sum_i P_i}{M_i / P_i}} + \frac{.15K_i}{\sum_i .15K_i} + \frac{T_i (T_i / R_i)}{\sum_i T_i (T_i / R_i)}$$

NOTE: M_i/P_i = per capital money income of i^{th} State (i.e., per capita income as defined by Census Bureau, not BEA).

T_i/R_i = "general tax effort factor" of i^{th} State as defined in the Act.

It can thus be seen that the 3-factor formula allocates funds to the States according to population, modified by the ratio of the State's "general tax effort" to its per capita money income. Since T_i/R_i is a numerator figure while M_i/P_i is a denominator figure under both formulas, the amount available to a State tends to be increased by a large State and local tax effort relative to the State's aggregate personal income, and tends to be reduced by a large per capita money income.

Allocation of funds to local governments

The Act divides each State into geographic areas called "county areas." (These are usually, but not necessarily, coterminous with the area of a county government. For example, the independent cities of Virginia would qualify as separate "county areas.") The Act provides that one-third of the funds allocated to a State shall be held by the State government for its own use, while two-thirds shall be allocated to the various units of local government in each "county area." The first step in this process is to determine the amount allocated to each "county area." "County area" is not a unit of government. It is an intermediate step. Additional notation is required.

A_{ij} = aggregated adjusted taxes of the j^{th} "county area" in the i^{th} State. Adjusted taxes constitute the aggregate taxes levied by all units of local government in the "county area", excluding those taxes allocable to education expenses. ("Units of local government" have a specific definition. In general they refer to units of general government. Special taxing authorities, e.g., Missouri fire protection districts, are not considered units of general government, and do not qualify for revenue sharing funds as described below.)

j = general subscript for a "country area."

Allocation to a "county area"

The relative amount allocated to the ij^{th} "county area" is given by the following formula:

$$\frac{\frac{P_i}{P_{ij}}}{\frac{M_{ij}/P_{ij}}{P_i}}$$

This formula is conceptually analogous to the 3-factor formula for allocation of funds to the States. It can be simplified to

$$\sum_j \frac{A_{ij}}{(M_{ij}/P_{ij})^2} = \frac{\text{aggregate "county area" adjusted taxes}}{(\text{"county area" per capita money income})^2}$$

$$\sum_j \frac{A_{ij}}{(M_{ij}/P_{ij})^2} = \sum_{\text{county areas}} \frac{\text{aggregate "county area" adjusted taxes}}{(\text{"county area" per capita money income})^2}$$

Allocation of "county area" funds among the units of local government in the "county area"

If in any "county area" there is an Indian tribe (or Alaskan native village) then initially "county area" funds are allocated to such tribes on the basis of the population of the tribe relative to the population of the "county area." The remainder of the "county area" funds are then allocated first to county governments, then to other units of local government.

Allocation to county government

k = general subscript for a county government. The relative amount allocated to the ijk^{th} county government is given by the formula

$$\frac{A_{ijk}}{A_{ij}} \cdot$$

Aggregate allocation to township governments

After allocation to county governments, the remaining amount is then reduced by an aggregate allocation to all township governments (if any) which are in the "county area."

Let A_{ijt} = aggregate adjusted taxes of all township governments in the "county area."

The relative aggregate amount set aside for all township governments is then given by the formula:

$$\frac{A_{ijt}}{A_{ij}}$$

This is an intermediate aggregate allocation. It does not specify the distribution of such allocation among the township governments.

Allocation to other units of local government (i.e., other than counties and townships)

After allocation of "county area" funds to county governments and aggregative township governments, the remaining funds are allocated to all other units of local government.

Let h = general subscript for a unit of local government (other than county or township).

The relative allocation to the ijh^{th} unit of local government is given by the following formula:

$$\frac{P_{ijh} \cdot \frac{A_{ijh}}{M_{ijh}} \cdot \frac{M_{ij}/P_{ij}}{M_{ijh}/P_{ijh}}}{\sum_h P_{ijh} \cdot \frac{A_{ijh}}{M_{ijh}} \cdot \frac{M_{ij}/P_{ij}}{M_{ijh}/P_{ijh}}}$$

This formula is analogous to the 3-factor formula for States.

As before, this formula can be simplified, and shown as

$$\frac{\frac{A_{ijh}}{(M_{ijh}/P_{ijh})^2}}{\sum_h \frac{A_{ijh}}{(M_{ijh}/P_{ijh})^2}} = \frac{\frac{\text{aggregate local adjusted taxes}}{(\text{local per capita money income})^2}}{\sum_{\text{local units}} \frac{\text{aggregate local adjusted taxes}}{(\text{local per capita money income})^2}}$$

Allocation of funds to specific township governments

The aggregate amount set aside for townships is distributed to the township governments by the same formula as that used for other units of local government. The subscript *h* then refers to a particular township government.

— — — — —

The allocations, including separate terms in the various formulas, are subject to additional constraints and conditions. For example, each term in the 5-factor formula has a specific aggregate dollar maximum, so that adjustments must be made before the terms can be added. There are other maximum and minimum conditions on the various allocations and terms. Some of these (but not all) are explained in general on the attachment, which is a copy of pages from the booklet, "What is General Revenue Sharing?", Office of Revenue Sharing, Department of the Treasury, August, 1973.

* * * * *

III. The Allocation Process

Several steps are taken to determine the amounts to be allocated to each state. The Office of Revenue Sharing distributes the \$6.05 billion allocated by Congress for the Fourth Entitlement Period (FY 1974) in the following manner:

1. \$6.05 billion is allocated among the states according to the 3-factor Senate formula (population, tax effort and income);
2. \$6.05 billion is also allocated among the states according to the 5-factor House formula (population, urbanized population, per capita income, state income tax collections, and tax effort);
3. The higher of the two amounts is selected for each state. Since the sum is greater than \$6.05 billion, each amount is scaled down proportionately so that the total allocation equals \$6.05 billion.
4. If either Alaska or Hawaii uses the 3-factor formula, its allocation is increased by the same percentage adjustment as applies to the base pay allowances of Federal government employees residing in those states (15 percent in Hawaii and 25 percent in Alaska).

The next step is to allocate *within* each state, according to the following process:

1. One-third of the state's allocation is paid to the state government, and the remaining two-thirds is apportioned to units of local government within the state.
2. The amount to be allocated to units of local government is divided by the population of the state to establish the per capita entitlement for all governments within the state.
3. The local government amount is distributed to county areas (these are geographic areas, not governments) based upon the ratio that each county area bears to all county areas within the state according to the formula: population X tax effort X relative income.
4. If this calculation allocates to any county area an amount which on a per capita basis, exceeds 145 percent of the per capita entitlement calculated in step 2., its payment is reduced to the 145 percent level and the resulting surplus is shared

proportionately by all the remaining unconstrained county areas within the state.

5. Similarly, if any county area is allocated less than 20 percent, on a per capita basis, of the amount calculated in step 2., its allocation is increased to the 20 percent level and the remaining deficit is taken proportionately from all the remaining unconstrained county areas within the state.

6. Each county area allocation is then divided into four parts:

- a. An amount for Indian tribal governments or Alaskan native villages is then determined on the ratio of tribal/village population to the total population of the county area.
- b. From the remainder, a township allocation is determined on the basis of the ratio of all township adjusted taxes to the total adjusted taxes in the county.
- c. A county government share is determined similarly, on the basis of county government adjusted taxes.
- d. The remaining proportion is for the other units of local government.

7. Townships and other local governments are then allocated funds on the basis of the formula: population X tax effort X relative income. If a unit of government receives more than 145 percent on a per capita basis, it is adjusted to the 145 percent level. If a unit receives less than 20 percent, its allocation is increased to the lower of either the 20 percent level, or 50 percent of its adjusted taxes and transfers. Then, if any unit receives more than 50 percent of its adjusted taxes and transfers, its allocation is reduced to that level and the excess is given to the county government.

8. If the county government has been allocated more than 50 percent of its adjusted taxes and transfers, its allocation is reduced to that level, and the excess is returned to the state government.

9. If any allocation is less than \$200, or any unit of local government waives its entitlement, those funds are allocated to the next higher level of government.

10. Finally, if the amounts allocated by the above procedure do not total 100 percent of the funds available for distribution, the appropriate adjustment is made to the entitlement figure in step 3. The process (step 3 through 8) is repeated until the amounts allocated total 100 percent of the funds available.

**Recent
Developments in
Estimation
for Local Areas**

Introduction

Paul S. Levy, University of Illinois

Recent Developments in Estimation for Local Areas—This session is sponsored jointly by the Committee on Small-Area Statistics and the Social Statistics Section of the American Statistical Association.

Statisticians and demographers working with health, economic, and social planners all are aware of the critical need for accurate small-area statistics, and have been frustrated by the rather primitive methodological tools available for obtaining estimates for local areas. It is only very recently that our survey statisticians and demographers have given some priority and have allocated resources to the development of some general methodology.

We are fortunate in having at this session two speakers who have been involved in very large methodological studies on small-area estimation. They are Maria E. Gonzalez of the Statistical Research Division, U.S. Bureau of the Census who has spent several years investigating the properties of a type of estimation process known as *synthetic estimation*, and Eugene P. Ericksen of the Institute of Survey Research at Temple University who has been developing and evaluating new methodology for estimating local populations. Our two discussants are Professor Richard M. Royall of the Department of Biostatistics, The Johns Hopkins University, who has written extensively on many aspects of sampling theory and Mr. Hyman B. Kaitz, now with C.S. & R. Associates and formerly Assistant Commissioner in charge of Current Employment Analysis of the Bureau of Labor Statistics.

Use and Evaluation of Synthetic Estimates

Maria Elena Gonzalez, U.S. Bureau of the Census

Definition of synthetic estimates

An unbiased estimate is obtained from a sample survey for a large area; when this estimate is used to derive estimates for subareas, on the assumption that the small areas have the same characteristics as the larger area, we identify these estimates as synthetic estimates. For the smaller areas, the estimates are no longer unbiased. However, it is possible to measure an average mean square error (MSE) for this set of estimates.

The simplest synthetic estimates are obtained by assuming that for the statistic of interest the mean value in the large area applies to each subarea directly; more refined estimates can be obtained by making this assumption for subgroups of the population. In the case when subgroups of the total are used, they should be nonoverlapping and exhaustive; the statistical estimates for the subgroups of the larger area are combined using independently known weights for the smaller area (e.g., as found at the time of the census) to obtain synthetic estimates for the smaller areas.

One is interested in estimating a characteristic, X . Identify j subgroups in the population, which are nonoverlapping and exhaustive. From the larger area we obtain estimates, x_j for $j=1, 2, \dots, G$.

A synthetic estimate is desired for subarea i , which is within the larger area. From the latest census we have weights P_{ij} , such that

$$\sum_{j=1}^G P_{ij} = 1.$$

The synthetic estimate, x_i^* , for characteristic X and subarea i is defined as

$$x_i^* = \sum_{j=1}^G P_{ij} x_j \quad (1)$$

This estimate associates the characteristic x_j of the larger area with each of the subareas i .

Use of synthetic estimates

Synthetic estimates are used primarily to develop small-area estimates when sample sizes are too small to give reliable results directly. Some examples of recent uses follow.

1) The National Center for Health Statistics has developed synthetic State estimates of disability based on the Health Interview Survey data. National rates of disability for 78 subgroups defined in terms of age, sex, size of household, income, industry, etc., were obtained from the data collected in the Health Interview Survey. These disability rates were weighted by the corresponding population in individual States, from the 1960 Census of Population, to derive synthetic State estimates of disability. [1]

2) The Bureau of the Census has used synthetic estimates for the imputation of population for units reported as vacant in the 1970 Census of Population and Housing, but which were actually occupied. A subsample of the housing units reported as vacant in the 1970 Census of Population and Housing was selected and interviewers were sent to these units to determine how accurately that determination had been made. About 11 percent of the housing units reported as vacant were determined to have been occupied at the time of the census. Separate estimates of such error rates were prepared for twelve geographic areas within the United States. Within each area the rate was applied to each enumeration district in the census and the applicable percentage of vacant units was converted to occupied units. The estimates of the error rates for areas such as cities, counties or States were synthetic estimates. [2]

3) In order to study the properties of synthetic estimates, an experiment was conducted to develop unbiased and synthetic estimates of unemployment for SMSA's for monthly, quarterly and annual estimates based on the Current Population Survey (CPS) data. A comparison of the reliability of the two types of estimates revealed that for monthly data the synthetic estimates were preferable, while for annual data the unbiased estimates were preferable; for the quarterly data, the two were of about equal reliability. [2]

4) In the 1960 Census of Housing enumerators were instructed to rate the physical condition of each housing unit into one of three categories: "sound," "deteriorating," or "dilapidated." One important purpose of this was to provide

data on substandard housing defined by Federal and local housing agencies as comprising units lacking complete plumbing facilities plus units which were dilapidated but had all plumbing facilities.

In the 1970 Census of Housing information was again obtained about plumbing, but synthetic methods were used to develop estimates of housing units which were dilapidated with all plumbing facilities (DWAPF). To obtain these estimates census data on housing units with all plumbing facilities were multiplied by estimated proportions of dilapidated housing units which had all plumbing facilities, as derived from a postcensus survey, Components of Inventory Change (CINCH).¹ From CINCH, estimates of DWAPF housing units were obtained for specified subgroups for 15 selected large SMSA's and for 4 balance of regions of the United States. Synthetic estimates for the smaller areas within these 19 geographic areas were derived using the corresponding set of DWAPF proportions.

Evaluation of synthetic estimates

Synthetic estimates are biased; to evaluate their reliability one can use the MSE, which can be expressed as the sum of the variance and the square of the bias:

$$MSE(x_i^*) = \sum_{j=1}^G P_{ij}^2 \sigma_{x \cdot j}^2 + (x_i^* - x_i)^2 \quad (2)$$

where:

- $\sigma_{x \cdot j}^2$ is the sampling variance of estimate $x_{\cdot j}$;
- x_i is the "true value" of the statistic for subarea i ; and
- x_i^* is the expected value of the synthetic estimate for subarea i .

The estimate given in formula 2 assumes that:

- a. the P_{ij} 's are fixed and measured without error; and
- b. the $cov(x_j, x_k) = 0$, for $j \neq k$.

In general, the values of x_i are not known and consequently the MSE of an individual synthetic estimate cannot be calculated for a particular area "i." However, if we establish M subareas within the survey population, the average MSE of the synthetic estimate over the M subareas (which may be unequal in size) can be estimated from the sample. Let

$$E \left[\frac{1}{M} \sum_{i=1}^M (x_i^* - x_i)^2 \right] = \alpha \quad (3)$$

The average MSE can be estimated by using the following approximation:

$$\hat{\alpha} \doteq \frac{1}{M} \sum_{i=1}^M (x_i^* - \sum_{j=1}^G P_{ij} x_{ij})^2 - \frac{1}{M} \sum_{i=1}^M \sum_{j=1}^G P_{ij}^2 (1 - 2f_{ij}) \sigma_{x_{ij}}^2 \quad [3] \quad (4)$$

where:

x_{ij} is the estimated statistic from the sample for subclass j and area i ,

f_{ij} is the sample estimate of the proportion of the total for the j -th subclass that is in the i -th subarea, that is

$$f_{ij} = n_{ij} / \sum_{i=1}^M n_{ij}, \text{ and}$$

$\sigma_{x_{ij}}^2$ is the sampling variance of estimate x_{ij} .

The interpretation of the square root of the mean square error is not altogether clear. Probability statements which can be made using the standard error for an unbiased estimate do not necessarily hold when the estimates are biased and the root mean square error is used as a measure of reliability. To try to understand the situation, an empirical study of the root mean square errors was made, using 1960 housing data. Synthetic estimates for a census were compared with actual measurements of the item for the same census to obtain an estimate of the bias. For various groupings of areas we then computed an average root mean square error (\overline{RMSE}); this estimate together with the distribution of the biases was then used to compare an empirical distribution of the biases of synthetic estimates with the normal distribution.

As part of the publication of the 1970 Census of Housing, data on housing units dilapidated with all plumbing facilities collected in the 1960 Census of Housing were available for comparison with a set of synthetic housing estimates of DWAPF derived from the same data. The average mean square error for a set of M areas is given by

$$\text{Average MSE} = \frac{1}{M} \sum_{i=1}^M (x_i - x_i^*)^2 \quad (5)$$

where:

x_i is the census estimate for area i .

The use of formula 5 to estimate the average MSE assumes that the second term of formula 4 is negligible. This assumption is reasonable for large areas. The square root of the average MSE gives the estimate for the \overline{RMSE} .

¹ The change in procedure in estimating DWAPF housing units was necessary because a majority of housing units in the 1970 Census of Housing were enumerated by a mail-out, mail-back procedure; in addition, studies of these data for 1960 indicated that statistics based on enumerator ratings are highly unreliable.

For all States we have two estimates of dilapidated housing units with all plumbing facilities in 1960; the 25-percent census estimate and a synthetic estimate based on a particular set of subgroups. The difference between these two estimates will be used as an estimate of the bias of the synthetic estimation procedure. Table 1 shows estimates of the proportion of the set of synthetic estimates for States with a relative bias within

specified values. The relative bias for an area is defined as the difference between the synthetic estimate and the census estimate divided by the synthetic estimate.²

²The synthetic estimate was used as denominator, instead of the reported estimate, because in 1970 the synthetic estimate was the only one available.

Table 1. Distribution of Relative Biases of Synthetic Estimates for States

State estimate	Number of areas	Proportion with relative biases				
		0-9%	10-19%	20-29%	30-49%	50%+
1,000-2,499.....	7	0.14	0.29	0.29	0.14	0.14
2,500-4,999.....	6	0.50	0.17	0.17	0.17	0.00
5,000-9,999.....	13	0.23	0.46	0.15	0.08	0.08
10,000-19,999....	16	0.38	0.38	0.19	0.06	0.00
20,000 or more...	8	0.38	0.38	0.25	0.00	0.00
Total.....	50	0.32	0.36	0.20	0.08	0.04

This table shows that the proportion of estimates with large relative biases diminishes as the size of the synthetic estimate increases. For example, for synthetic estimates between 1,000 and 2,499 DWAPF housing units, 57 percent have relative biases of at least 20 percent; however, for synthetic estimates of over 10,000 units only about 25 percent have relative biases greater than 20 percent. When we consider the State synthetic estimates for all States, we note that 32 percent have relative biases of 20 percent or more, 12 percent have relative biases of 30 percent or more and 4 percent have relative biases of 50 percent or more. The average number of DWAPF housing units for States is about 13,000; the estimated average root mean square error is about 2,500; the ratio of the \overline{RMSE} divided by the average size of State synthetic estimates of DWAPF housing units is 0.19. A

high variability of the synthetic estimate is shown by the fact that the \overline{RMSE} divided by the mean is about 20 percent. This shows that the synthetic estimates obtained do not account for a large part of the variability among areas. The synthetic estimates of housing units DWAPF are computed using a particular set of subgroups, defined in terms of tenure, race of head of household and other characteristics related to the quality of housing unit. The use of other subgroups would produce a different set of synthetic estimates.

From the point of view of ascertaining whether the average root mean square error can be used to make probability statements the results are more encouraging. Table 2 gives some

Table 2. Comparison of Empirical Distribution of the Biases of State Synthetic Estimates of Dilapidated Housing Units with All Plumbing Facilities with the Theoretical Normal Distribution

Multiple of standard error (σ)	Normal probability	Distribution of bias of State estimates (n=50)		
		Total $\overline{Aver} = 12,970$ $\overline{RMSE} = 2,490$	Inside SMSA's $\overline{Aver} = 8,170$ $\overline{RMSE} = 1,540$	Outside SMSA's $\overline{Aver} = 4,800$ $\overline{RMSE} = 1,340$
0.50	38.0%	48%	50%	62%
0.75	55.0	62	62	68
1.00	68.0	74	68	74
1.25	79.0	86	82	84
1.50	87.0	88	90	88
1.75	92.0	88	90	88
2.00	95.0	92	94	94
2.25	97.6	94	94	98
2.50	98.8	96	96	98
3.00	99.7	100	100	98

comparisons of the distribution of the difference between State synthetic estimates of DWAPF housing units for 1960 and the estimates reported in the 1960 census. The first two columns of the table show the expected percentage of the normal distribution at different multiples of the standard error (σ). For example, 95 percent of the normal distribution is expected within two standard errors of the mean. The empirical distributions of the biases of the synthetic State estimates of DWAPF housing units are given in columns 3, 4, and 5. For example, 48 percent of the biases for estimates of total for States are less than one-half the estimated RMSE.

The empirical distributions of the biases of synthetic State estimates are closer to the mean (on the average) for values within one standard error of the mean, than expected for the normal distribution. For example, for half a standard error the normal distribution expects to cover about 38 percent of the distribution; for State totals, the empirical distribution actually includes 48 percent of the distribution; for estimated units within SMSA's, the empirical distribution includes 50 percent of the distribution and for units outside SMSA's the empirical distribution includes 62 percent of the distribution. However, for values which are more than two standard errors from the mean, the empirical results are reversed: the frequency of synthetic estimates with biases more than two standard errors

from the mean is greater than expected for normal distributions; for State synthetic estimates about 8 percent had biases which differed by more than two standard errors from the mean. That is, on the average there are more outliers for synthetic estimates than would be expected for a normal distribution.

Table 3 shows empirical distributions of the biases of the estimates of DWAPF housing units for non-Negro renters in counties within SMSA's. The distribution was computed separately depending on the magnitude of the estimate of DWAPF housing: less than 100, 100 to 499, and 500 or more units. It is possible to carry out the analysis separately for the three groups through the use of formula 5 for the average root mean square error. The results reveal a similar pattern to the results for State estimates given in table 2. For values within one standard error the empirical distribution gives conservative estimates of the probability of occurrence, except for the distribution of DWAPF housing units with 500 or more units. However, for values at three standard errors we find more outliers than expected for the normal distribution. For example, for synthetic DWAPF estimates less than 100, the normal distribution expects only 0.3 percent of the cases to be further than three standard errors, but we find that 2.5 percent of the values have biases larger than three times the average root mean square error.

Table 3. Comparison of Empirical Distribution of the Biases of Synthetic Estimates for Non-Negro Renters in Counties within SMSA's of Dilapidated Housing Units with All Plumbing Facilities with the Theoretical Normal Distribution

Multiple of standard error (σ)	Normal probability	Distribution of bias for non-Negro renters in counties within SMSA's		
		DWAPF <100 RMSE = 39 n = 160	DWAPF (100-499) RMSE = 125 n = 219	DWAPF (500+) RMSE = 300 n = 84
0.50	38.0%	50.0%	49.0%	37.0%
0.75	55.0	66.0	66.0	52.0
1.00	68.0	84.0	79.0	64.0
1.25	79.0	89.0	85.0	75.0
1.50	87.0	93.0	89.0	85.0
1.75	92.0	94.3	92.2	89.2
2.00	95.0	95.6	94.5	95.2
2.25	97.6	96.2	96.8	98.8
2.50	98.8	96.2	97.7	98.8
3.00	99.7	97.5	98.6	98.8

Conclusions

Census data allow us to compute synthetic estimates and to compare them directly to the census estimates. Therefore, the biases of synthetic estimates can be obtained and their distribution analyzed directly.

The results presented comparing 1960 estimates of dilapidated housing units with all plumbing facilities with synthetically derived estimates show that the synthetic estimates are highly variable, but that the distribution of their biases is not too far from normal.

The analysis presented is based on a particular set of synthetic estimates; alternative sets using other variables should be investigated in order to be able to select the subgroups which account for a large proportion of the variability of the local area estimates, with an aim toward improving local area estimates. The results presented here, based on a particular set of synthetic

estimates, may not necessarily generalize to possible alternative sets of synthetic estimates.

References

[1] "Synthetic Estimates of Disability," published in 1968 by the National Center for Health Statistics, PHS publication No. 1759.

[2] A more detailed discussion of this project is available in "Estimation of the Error of Synthetic Estimates," by Maria Elena Gonzalez and Joseph Waksberg, presented at the first meeting of the International Association of Survey Statisticians, Vienna, Austria, August 18-25, 1973.

[3] Appendix 1 of "Estimation of the Error of Synthetic Estimates," by Maria Elena Gonzalez and Joseph Waksberg gives the derivation of the approximation given in formula 4.

Developments in Statistical Estimation for Local Areas

Eugene P. Ericksen, Temple University

Introduction

The regression-sample data method of postcensal estimation is a procedure by which one can combine sample survey data with symptomatic information to obtain local estimates of the criterion variable being measured by the survey data. This method has been tested extensively using population growth as the criterion (Ericksen, 1973a, 1973b), first, for the period beginning in 1960 and ending in 1964 to 1967, and then more extensively for 1960 through 1970. The steps of the procedure in the latter test were as follows:

- a. Sample estimates of population growth were obtained for the primary sampling units selected into the national sample of the Current Population Survey. These 1970 estimates of current population were divided by the corresponding 1960 census populations giving sample estimates of 1960 to 1970 population growth.
- b. Symptomatic indicators, in this case 1970/1960 ratios of births, deaths, and school enrollment, were compiled for the sample psu's (primary sampling units) and a multiple regression equation was computed using the sample estimates of population growth as the depen-

dent variable. A second equation was then computed using the series of ratio-correlation estimates calculated at the Population Division of the Bureau of the Census as a fourth symptomatic indicator.

- c. Values of the symptomatic indicators for counties were substituted into the regression equations and estimates were made of the 1960 to 1970 population growth. This step was carried out for 2,586 counties in 42 States.

Corresponding estimates for these counties were made at the Population Division using four standard demographic techniques which have traditionally been used to estimate population growth. Of these techniques, the ratio-correlation technique was the most accurate. Little was gained from averaging estimates of two or more standard techniques. The regression estimates produced by our combination of sample data and symptomatic information were more accurate than those of any single or combination of standard techniques. This was particularly true when the series of ratio-correlation estimates was added as a fourth symptomatic indicator. There were moderate reductions in the mean error, but the greatest gain was in the reduction of the number of large errors, which was over 20 percent. These results are presented in table 1.

Table 1. Relative Accuracy of Standard and Regression-Sample Data Estimates of Population Growth, 1960 to 1970, for 2,586 Counties in the United States

Procedure	Mean error ¹	Number of counties with error 10 percent or greater
Vital Rates	7.4	673
Component Method II	7.2	645
Composite	5.9	407
Ratio-Correlation	4.6	264
Component Method II, Composite, Ratio-Correlation, Averaged ²	4.7	249
Regression-Sample Data, Three Symptomatic Indicators ³	4.4	220
Regression-Sample Data, Four Symptomatic Indicators ³	4.2	194

¹ All estimates were multiplied by an appropriate constant in order to sum to a separately estimated 42-State total.

² This was the most accurate combination of the four standard techniques.

³ The 3-variable equation used was $\hat{Y} = .158 + .218X_1 + .142X_2 + .520X_3$. The 4-variable equation used was $\hat{Y} = .058 - .097X_1 + .045X_2 + .214X_3 + .745X_4$.

Source of Standard Estimates and Estimate of 42-State Total: U.S. Bureau of the Census, Current Population Reports, Series P-26, No. 21, "Federal-State Cooperative Program for Local Population Estimates: Test Results - April 1, 1970," Washington, D.C.: Government Printing Office, 1973.

Some of the prominent features of our new method of postcensal estimates are the following:

- a. Estimates of population growth have been shown to be more accurate. Part of the reason for this gain is that it is not necessary to make any assumptions concerning the nature of relationships beyond those of least squares linear regression. One of the difficulties of the ratio-correlation technique, for example, is the assumption of the continuance of past relationships.
- b. Other series of estimates can be incorporated as symptomatic indicators. By including the series of ratio-correlation estimates as a symptomatic indicator, we have a way of correcting for the bias of ratio-correlation which arises from assuming the continuance of past relationships.
- c. There is a procedure by which the mean squared error of the regression estimates can be calculated. Given this facility for measuring error, we can systematically test various combinations of symptomatic indicators to determine the composition of the optimal set. Because of the presence of the within-psu sampling error, this does not necessarily include all available symptomatic indicators.

Current Activity at the Bureau of the Census

A project is currently under way at the Bureau of the Census to compute yearly estimates of population growth since 1970 by our regression-sample data method. A determined effort is being made to obtain symptomatic data for all counties in each of the 50 States and the District of Columbia. It now appears that births, deaths, and school enrollment will be available, but with some time lag, for counties in all but a small handful of States. Additional data on automobile registrations will be available for some States and it is also expected that data on income tax exemptions will be available for all States (Zitter and Word, 1973). Substantial gains are anticipated from the use of tax records, even outside our regression-sample data format. In view of the changing relationships among variables, and the possibility that other symptomatic indicators will become available, the following instructions are pertinent to potential users of the regression-sample data technique:

- a. Applications of and experimentation with the ratio-correlation technique have shown conclusively that relationships among a given set of variables can be expected to change over time. We have shown that the series of regression-sample data estimates were relatively accurate when computed over a given 10-year period for particular sets of three and four symptomatic variables. However, the accuracy of the regression-sample estimates relative to those of other techniques could change for a shorter estimation period beginning in 1970. It is also possible that the most accurate regression-sample data estimates would be computed with a different set of symptomatic indicators in this period. We can test this possibility by inspecting the mean squared error of the regression

estimates and the correlations of the various indicators with the sample estimates of population growth.

- b. In the absence of correlations between the sampling error and the value of the symptomatic indicators, the estimated regression equation using the sample estimates of population growth as the dependent variable is an unbiased estimate of the regression equation which would be obtained using census tabulations of population growth if they were available. However, the presence of the within-psu sampling error will lower the observed values of the correlation coefficients. Low observed values of the correlation coefficients do not necessarily mean that the errors of the regression estimates will be large.
- c. There are some unsolved problems regarding the inference from a sample of psu's to a universe of counties. The mean squared error of the regression estimates refers to the accuracy of estimates for psu's when the units of interest may be counties. To the extent that counties are different from psu's, reductions in the mean squared error for psu's may not improve the accuracy of estimates for counties. A second unresolved problem has to do with specification errors arising from the distribution of the within-psu sampling errors. If the size and direction of these errors vary systematically with values of the symptomatic indicators, the assumptions of linear regression may not be met. We have found this to be a minor problem in our application that resulted in larger errors for units with extreme growth rates, but the problem could be more important in other applications.

The Mean Squared Error

We have shown elsewhere (Ericksen, 1973a, 1973b) that the mean squared error of the regression-sample data estimates can be expressed by the formula:

$$\frac{E(Y - \hat{Y})^2}{n} = \frac{(n - p - 1)\sigma_u^2}{n} + \frac{(p + 1)\sigma_v^2}{n} \quad (3.1)$$

where:

- σ_u^2 = the between-psu variance unexplained by the indicators,
- σ_v^2 = the within-psu variance,
- n = the number of psu's in the sample, and
- p = the number of symptomatic indicators.

When n is large relative to p , the mean squared error is determined by (1) σ_u^2 , which decreases when new symptomatic indicators are added, and (2) the within-psu component of error which increases when indicators are added. If there were no within-psu component of error, optimal results would be obtained by maximizing p , i.e., by utilizing all available symptomatic information. We have found in our applications, however, that the presence of within-psu sampling variability has often meant that the optimal set of symptomatic indicators did not include all that were available.

In the test of 2,586 counties, there were seven symptomatic indicators available: births, deaths, school enrollment, and the four standard estimates. As shown in table 2, where the ratios of the 1970 to the 1960 decennial census populations were the dependent variable, gains in the accuracy of regression estimates for psu's were obtained by increasing the number of symptomatic indicators from four to seven. However, in the more

realistic application, when the within-psu component of error was present, the increase from four to seven indicators actually brought about an increase in the errors. The mean error of the 2,586 county estimates increased from 4.2 percent to 4.7 percent. A similar result was obtained when 6 variables, with 51 observations (1 for each State and the District of Columbia) were available.

Table 2. Mean Errors Obtained With Various Sets of Symptomatic Indicators

Number of symptomatic indicators ²	Mean Percentage Error ¹	
	Dependent variable 1970 census/1960 census	Dependent variable 1970 CPS/1960 census
Units of estimates are 444 primary sampling units		
3 ²	2.83	3.20
4 ²	2.60	2.92
7 ²	2.11	3.24
Units of estimates are 50 States and District of Columbia		
3 ³	1.22	1.64
4 ³	1.08	2.16
6 ³	1.07	3.91

¹Mean percentage error, comparing regression estimate with Census Bureau tabulation. "Dependent variable" is that used to compute regression equation.

²Set of three indicators included births, deaths, and school enrollment. The fourth indicator was the ratio-correlation estimate, and indicators five through seven were the composite, component method II, and vital rates estimates.

³Set of three indicators included births, school enrollment, and work force. The fourth indicator was deaths, and indicators five and six were automobile registrations and income tax returns.

The fact that the optimal set of indicators included four variables was due to the nature of the structural relationships and the size of the within-psu variance. We have evidence that these change over time, as shown in table 3. In particular, for the Current Population Survey (CPS) sample, the within-psu variance increased. This is because the CPS sample was based on the 1960 census, and that patterns of subsequent growth were uneven, leading to variation in the size of sample segments within psu's. This trend leads us to expect that more symptomatic indicators should be used in shorter time periods. On the other hand, the relationships among the variables appear to become stronger as time passes. In spite of the increasing within-psu variability which dampens the observed correlations, these observed correlations grew larger from 1963 through 1967. In shorter periods, changes in population size, as well as in the symptomatic indicators, appear to be due more to random fluctuations. As times passes, changes in the variables are larger, and the relationships among these changes more systematic. This leads to the contrary expectation that the

optimal set of indicators would be smaller for a shorter time period. To determine the optimal set of indicators, we must estimate the mean squared error in each estimating situation.

Because the true values of the criterion variable are unobserved, the mean squared error cannot be estimated directly. To obtain equation (3.1), we must first compute the mean of the squared differences between the regression estimates and the sample estimates for the sample psu's and then subtract an allowance for the within-psu sampling error. The mean squared difference between the regression and sample estimates can be expressed by the formula:

$$\frac{E(Y_o - Y)^2}{n} = \frac{(n - p - 1)}{n} (\sigma_u^2 + \sigma_v^2) \quad (3.2)$$

To obtain (3.1) we need to subtract the term $(n - 2p - 2)\sigma_v^2/n$. In order to obtain a good estimate of the mean squared error, we clearly need to have a good estimate of σ_v^2 .

Table 3. Values of Estimated Within-Psu Variance of Population Growth and Coefficients of Determination, 1963 through 1967

Year	Within-psu variance ¹	Coefficient of determination (R^2) ²
1963	.0253	.016
1964	.0378	.021
1965	.0383	.085
1966	.0458	.117
1967	.0473	.264

¹Computed as squared difference between random half-samples defined by rotation group.

²Three symptomatic indicators were births, deaths, and school enrollment in each case.

When we reported earlier results (Ericksen, 1973b), we did not feel that a good estimate of σ_v^2 was available. We had computed half-samples defined by the eight rotation groups of the CPS (U.S. Bureau of the Census, 1963) and had overestimated the mean error of the sample estimates for psu's. This is because sample segments within the CPS sample had not been placed equally into rotation groups within individual psu's. However, when the half-samples were formed on the basis of sample segments without regard to rotation group, a better estimate was obtained. The derivation of equation (3.2) depends on the values of (1) the sampling error and (2) the structural errors of regression, along with the sampling errors being unrelated to the symptomatic indicators. Our technique for estimating the mean squared error is particularly sensitive to these specification errors, as the following illustration shows.

The practical question we faced in the 1970 test was whether or not improvements in accuracy over that given by the ratio-correlation technique would be obtained by adding births, deaths, and school enrollment as symptomatic indicators in a regression equation. We found that the ratio-correlation estimates accounted for 92.7 percent of the variance of the actual 1970/1960 ratios of population of the sample psu's. Adding the three symptomatic indicators, the coefficient of determination, R^2 was increased to .951, a clear increase in the explained and reduction in the unexplained variance. However, the increase in the explained variance of the sample estimates of 1960 to 1970 population growth obtained by adding the three symptomatic indicators to ratio-correlation was much smaller, from 41.7 percent to 43.0 percent. This was due to the presence of the within-psu error which is not reduced by adding symptomatic information. The observed variance of the distribution of sample estimates before regression was .0438. Using the series of ratio-correlation estimates as a single symptomatic indicator, the mean squared difference of the regression and sample estimates as expressed by equation (3.2) was .0255. This was reduced to .0250 when the number of symptomatic indicators was increased from one to four. Our estimate of the within-psu variance is $\sigma_v^2 = .0253$. Subtracting the allowance for this component of error, our final estimates of the mean squared error are .0004 where the ratio-correlation estimate is a single indicator and .0001 with four indicators. This is a very small difference considering the size of the within-psu variance and the mean squared difference between the regression and sample estimates. A small fluctuation could have seriously altered the observed results. When the number of symptomatic indicators was increased to seven, the coefficient of determination was R^2

= .432, the mean of the squared differences was .0249, and the final estimate of the mean squared error, .0002. These differences are so small that one may be on safer, although less scientific, grounds simply to observe that the increase in R^2 from .417 to .428 is large enough to produce a good reduction in error while guessing that the further increase to $R^2 = .432$ is not, given the increase in the number of symptomatic indicators.

Some of the difficulties in estimating σ_v^2 arise from the fact that the within-psu sampling error is positively correlated to the growth rate, and hence to the values of the symptomatic indicators. The correlation between the actual error of the CPS estimate and the estimated within-psu variance is +.45. This affects the estimate both of σ_v^2 and the way we obtain an estimate of equation (3.1) from equation (3.2). A second source of error is the correlation between the within-psu error and the growth rate, which is +.06. This introduces curvilinearity, since the sample estimates of the fastest growing areas tend to be too large and those of the slowest growing too small, thus biasing the estimation of regression coefficients. One result of this was that estimates of areas with extreme growth rates had larger errors. This particular problem is covered in the literature on econometrics where the usual solution is to apply a transformation. We have attempted several such solutions, but have yet to find a transformation which allows us to reduce the errors of the extreme cases without increasing the errors of the majority of cases which have moderate values.

One obvious procedure for reducing the mean squared error of the regression estimates is to reduce the within-psu variance. This could be done by improving the within-psu sample design, or, as we will attempt to do, by introducing more sample data. In our program at the Census Bureau, we plan to eventually request tabulations from other government surveys such as the Health Interview Survey and the National Crime Survey. This will reduce σ_v^2 where the psu's in the various surveys are the same and reduce the ratio $(p + 1)/n$ and therefore the within-psu component of error in equation (3.1) in cases where the psu's are different.

This necessarily reduces the errors of the primary sampling units, but the effects on county estimates are uncertain. To illustrate this point, when the regression equation with three symptomatic indicators, births, deaths, and school enrollment, was recomputed using the 1970/1960 census population ratios as the dependent variable, i.e., setting σ_v^2 equal to zero, the

mean error of the psu estimates was 2.8 percent. This compares to the mean error of 3.2 percent when the CPS estimates were the dependent variable. The difference between 2.8 and 3.2 percent was due to the within-psu error. However, when the two equations were used to make county estimates, the mean error was 4.4 percent in both cases. The Census Bureau's ratio equation, computed without the within-psu error, had done a better job of making psu estimates, but the transition from psu's to counties had become more difficult. When the distribution of errors was broken down by size of the 1970 county population, it was found that use of the decennial census ratios in place of the CPS estimates had reduced the mean error for all categories of counties with population greater than 25,000, but that the mean error had increased among counties smaller than 25,000. Counties in this last category were the majority of all counties but were least similar to the CPS sample psu's which usually consisted of combinations of counties picked with probabilities proportional to the size of the total population.

New Strategies and Plans

Given the limited gains obtained from reducing the within-psu component of error, and our lack of success in finding suitable transformations to reduce errors, the most promising approach to reducing our errors appears to be the introduction of new symptomatic information. One variable which has been shown to reduce errors is automobile registrations. Data were available in the 1970 test for 2,223 counties in 32 States. A five-variable regression equation, also including births, deaths, school enrollment, and the ratio-correlation estimate was computed and county estimates made. The mean error of these estimates was 3.8 percent and 122 errors were greater than 10

percent. The corresponding figures for this set of counties for the four-variable regression equation omitting automobile registrations were 4.1 percent with 148 large errors and, for the standard series of ratio-correlation estimates, 4.5 percent with 220 large errors.

Another promising, but as yet untested, variable is the number of exemptions on income tax returns. Changes in address of persons listed on income tax forms are to be used to estimate net migration and when added to recorded natural increase, could give extremely accurate estimates of population growth. It is quite possible that these estimates would be sufficiently accurate in themselves so that little gain would be obtained by computing regression-sample data estimates. But it is more likely that some bias will be introduced because of the characteristics of persons not listed on tax forms or whose likelihood of being listed on a form varies at point of origin and destination. In such a case, this bias could be corrected by using the tax estimate as a symptomatic indicator in a regression equation possibly including other symptomatic indicators with sample data as the dependent variable.

Finally, we have made plans to attempt to estimate other variables such as racial composition, unemployment, and median family income. Although births and deaths are available by race in many counties, the chief barrier faced here is the lack of symptomatic information. Data on wages and work force appear to be available in metropolitan areas, but we are still searching for symptomatic data available on a national basis. If such data can be found, we can combine our symptomatic information and sample data with estimates which can be generated by other means. One such series would be the synthetic estimates being discussed in this session.

Bibliography

- Ericksen, Eugene P. "A Method for Combining Sample Survey Data and Symptomatic Indicators to Obtain Population Estimates for Local Areas." *Demography* 10 (1973a): 137-160.
- Ericksen, Eugene P. "A Regression Method for Estimating Population Changes of Local Areas." 1973b, in press.
- U.S. Bureau of the Census. *The Current Population Survey--A Report on Methodology*. Technical Report no. 7. Washington, D.C.: U.S. Government Printing Office, 1963.
- Zitter, Meyer, and David Word. "Use of Administrative Records for Small-Area Population Estimates." Presented at the Annual Meeting of the Population Association of America, 1973.

Acknowledgment

The research upon which this paper is based was carried out in cooperation with the Bureau of the Census, in particular with Benjamin J. Tepping of the Research Center for Measurement Methods and with the Population Division, who supplied all the symptomatic information used in computing the regression equations as well as special tabulations concerning the distributions of errors of the ratio-correlation and other standard estimates. Special thanks are due to Lori Kessler and Harris Miller without whose assistance at Temple University this research could not have been carried out. The findings, recommendations, and conclusions in this paper are the sole responsibility of the author and are not necessarily endorsed by the U.S. Government. The data in this paper are the result of tax-supported research and, as such, are not copyrightable. The data may be freely reprinted with the customary crediting of the source.

Discussion of Papers by Gonzalez and Ericksen

Richard Royall
The Johns Hopkins University

Both of these papers are good examples of the process of developing estimators using conventional finite population sampling theory. We can pick out three important stages in this process:

1. Assuming observations on certain variables are available, scratch your head and write down an estimate which has some intuitive appeal.
2. Try to get a handle on bias and variance. (Having done this a few times and produced a few estimates, compare their mean square errors. Find one estimate that is better than another under certain assumptions about population parameters.)
3. Get a real population and try out the estimates to see which works better under various realistic conditions.

After, or along with, these three basic steps comes the secondary problem of measuring the uncertainty in an estimate. This usually boils down to finding a nearly-unbiased estimate of an approximation to the variance or mean square error. Unfortunately, these variance estimates rarely have the "face validity" or obvious reasonableness of the original statistic. For example, the synthetic estimates are in a gross sense reasonable. They obviously won't give really precise estimates, but they will be, if not in the right ballpark, at least in the right city. The variance estimate, on the other hand, might not even be on the right planet—a negative variance estimate might be reasonably described as "lost in space."

I would like to see a different approach used, and I think the problem at hand, estimation for small areas, is one in which this approach would yield different and better results than the conventional one, particularly with regard to providing estimates of mean square errors to use as measures of uncertainty. This approach would begin not with an estimate, but with an attempt to express the basic relationships among the relevant variables through a probabilistic model. The model would then be used to generate estimates, provide a framework for comparing estimates, and to provide estimates of standard errors. Often the conventional intuitive estimates are optimal or nearly so under a simple probability model, but sometimes the model suggests practical improvements, especially in the conventional measures of uncertainty. Varying the model can give valuable insight into the robustness of estimators. This general approach has been called "the prediction approach" because, when viewed in the context of (super-population) probability

models, many finite population inference problems are mathematically equivalent to classical prediction problems. "The prediction approach" actually has many facets—simple linear least-squares, esoteric fiducial, and full-blown Bayesian prediction techniques are only some of those available.

What would be the results of applying the prediction approach (least-squares variety) to the present problem? Two important general results I would expect are:

1. New estimators and new variance estimators for the old ones.
2. New insight into relationships among estimates already proposed and increased understanding of their strengths and weaknesses.

Specifically . . . I don't know what results would be obtained. The work has not, to my knowledge, been done. But some relevant comments can be made.

The "ratio-correlation" method and the "regression-sample data" method aren't so much two different methods as two different estimates, each more or less appropriate under its own prediction model. Although the two models do employ slightly different functions of births, etc. as regressors, the most important differences between these two estimates come not from different assumptions concerning the relationships among the relevant variables, but from different assumptions about available data. The ratio-correlation method is not allowed to use the sample data, while the regression method employs only data from the sample and the most recent census, ignoring the previous census. In both models, the total for a local area at one time is represented as a multiple of the total at an earlier time plus an error whose variance is proportional to the square of the earlier total. (We might ask whether a different error-variance might be more appropriate. If it is, this would suggest different estimates.) The multiplicative factor for a given area is a function of various bits of data concerning births, deaths, number of school children, etc. in that area. In this factor are certain coefficients which change over time. The "ratio-correlation method" uses estimates of out-of-date coefficients, while the "regression-sample-data" method uses less precise estimates of more timely coefficients.

When the "ratio-correlation" estimate is used as a "symptomatic indicator" in the "regression-sample-data" estimate, we are, in effect, using a particular linear combination of estimates

of the "old" coefficients and the "new." I think a formal model, in which coefficients for one time interval are stochastically related to those for an earlier interval, would be quite useful in evaluating this and other estimates based on all the data, from both censuses as well as the sample.

In much the same way, the choice between direct estimation and imputation in the synthetic estimation paper is really the choice between a high-variance estimate of a directly relevant parameter and a low-variance estimate of a different quantity. The choice need not be made—surely a combination of the two

is better than either taken alone. A probability model can express the relationships whose existence makes the whole notion of "imputation" reasonable. Such a model would generate (via standard linear prediction techniques) statistics which would give proper weight to both direct and imputed estimates.

I think, however, that one of the possibilities suggested by Gonzales and Waksberg in their Vienna paper is more promising—before really good local area estimates are produced, the synthetic estimation approach must move towards Ericksen's in making greater use of available local area variables.

Discussion of Papers by Gonzalez and Ericksen

Hyman B. Kaitz
CSR Associates, Alexandria, Va.

Both of these papers represent progress in the development of procedures for estimating local area data. Nevertheless it is work in progress; we clearly have some distance to go in attaining fully acceptable methodologies.

Both papers examine the accuracy of their techniques in terms of mean square errors of all the local area estimates, and in terms of the percentage of areas whose estimates deviate by more than a certain percentage, say 10 percent, from the known criterion values. It is desirable and relevant, of course, that such measures of dispersion or accuracy, be used in judging these various techniques.

There is a particular problem, however, which arises in this connection. Estimates are produced for areas which are individually identified. For example, the method may yield a specific estimate for Altoona. Is Altoona interested in knowing whether its estimate comes from a body of area estimates with a satisfactory mean square error, or is it interested in knowing how good the Altoona estimate is? I think the latter is more likely to be the case. In the absence of any alternative estimates, Altoona can perhaps become reconciled to its estimate, but it may not. It is interesting in this connection to see what has happened to local area population counts from the 1970 census. Even though these are presumably "complete" and official government counts, they have been subjected to strong criticism by various local interest groups who have said that the presumed population undercount should be officially allocated to local areas differentially (black-interest groups), or that the Census Bureau should use more accurate enumeration techniques for the Spanish-Americans.

When the data at issue are statistical estimates rather than official counts, there may even be more opportunity for criticism. For example, local area unemployment rate estimates have been subjected to criticism for a number of years in States like California, Ohio, New Jersey, and Massachusetts, principally because these estimates based on administrative records from the unemployment insurance system could be compared with the presumably comparable rates for the same areas based on the Current Population Survey. The prospect for critical examination of local area estimates is particularly strong when concrete incentives are present to seek estimates which increase local allocation of government funds under programs such as revenue sharing.

This suggests an approach to local area estimation which is not based on the use of a single technique, but on all the information available for a given locality which may help to improve its estimate. This would be a highly professional-labor-intensive activity and would probably be out of reach of most organizations with limited budgets.

With respect to the Ericksen paper a question may be raised about the possible use of additional local area information available generally, which may be used as symptomatic variables, such as:

1. the racial mix in the base year
2. the urban-rural mix in the base year
3. the population density in the base year
4. the age-sex mix in the base year, and so on.

Ericksen's work uses ratios as the symptomatic variables as well as the criterion variable. I would like to see some evidence on the amount of collinearity present among the symptomatic variables. In general, on statistical grounds, one seeks for symptomatic variables which are relatively uncorrelated with each other but are correlated with the criterion variable. These considerations should not ignore those based on the subject matter under study.

I liked Ericksen's formulation of the regression equation to take account of the sampling error in the criterion variable. This assumes independence of the u and v error terms which does not appear to be quite true for his data set. This suggests extension of his model to include some nonzero but unspecified covariance between u and v . In addition there may be reason to assume a v term for specific symptomatic variables, which would alter the model somewhat and would appear to produce biased estimates of the parameters of the regression equation. There is also some heterogeneity in u , which suggests seeking some transformation of the variables to correct this, or the use of generalized least squares in the estimation procedure.

Between the extreme of seeking a single estimation equation for all areas, and the extreme of seeking to maximize the use of ad hoc local area information through a variety of techniques, there may be an optimum point at which clusters of areas may be studied, each cluster with its own estimation technique. Ericksen's formulation suggests little payoff here if the analysis uses sample-based criterion variables.



DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE
OFFICE OF THE SECRETARY
WASHINGTON, D.C. 20201

Office of Human Development
Administration on Aging

TECHNICAL ASSISTANCE MEMORANDUM

AoA-TA-77-4

February 16, 1977

TO : STATE AGENCIES ADMINISTERING PLANS UNDER TITLES III AND VII
OF THE OLDER AMERICANS ACT OF 1965, AS AMENDED

INFO FOR: Area Agencies on Aging and Nutrition Projects

SUBJECT: Provisions of the State and Local Fiscal Assistance (General
Revenue Sharing) Act Amendments of 1976 (P.L. 94-488)

CONTENT: Public Law 94-488 (October 13, 1976), "State and Local Fiscal
Assistance Amendments of 1976," contains several provisions
that will affect the elderly.

1. Background

The State and Local Fiscal Assistance Act of 1972, commonly referred to as "General Revenue Sharing" (P.L. 92-512), was enacted on October 20, 1972. The legislation created a trust fund to receive a portion of Federal individual income taxes for disbursement to States and local units of government to be used for various types of programs, including social services for the poor or aged. Funds amounting to some \$6 billion annually were disbursed to approximately 39,000 recipient government entities. The program was scheduled to expire on December 31, 1976.

Public Law 94-488 was enacted on October 13, 1976 to amend and extend the revenue sharing effort.

2. Extension of Program

The 1976 Amendments extend the program from January 1, 1977 through September 30, 1980. The new legislation contains several provisions that affect the elderly.

3. Major Provisions of Public Law 92-512 Affecting the Elderly, and Effect of Amendments thereto:

A. Use of Funds by Local Governments for Priority Expenditures

The original Act provided that revenue sharing funds received by units of local government could be used for priority expenditures covering ordinary and necessary maintenance and operating expenses, and capital projects for, among others, social services for the poor or aged. The Amendments eliminate the list of priority expenditures. Nevertheless revenue sharing funds can still be used to support programs which will benefit the elderly. This memorandum will offer a number of suggestions about how revenue sharing can be obtained for such programs.

B. Use of Funds for Matching

The Amendments specify that revenue sharing funds now may be used to match Federal grants under other Federal programs, such as those authorized by the Older Americans Act under Titles III (State and Community Services), V (Multipurpose Senior Centers), and VII (Nutrition Program for the Elderly). Hence, State and local agencies should make concerted efforts to secure revenue sharing support for their programs, both as a means of obtaining additional funds and as a means of satisfying non-Federal matching requirements. For more on how to pursue these objectives, see Section 4, below.

C. Nondiscrimination Provisions

The Amendments broaden the existing nondiscrimination provision in the original Revenue Sharing Act by adding

"Any prohibition against discrimination on the basis of age under the Age Discrimination Act of 1975 ... shall also apply to any such program or activity."

The Amendment to prohibit discrimination on account of age will not take effect until the Age Discrimination Act of 1975 (P.L. 94-135, dated November 28, 1975) becomes effective to prohibit age discrimination in all Federal programs, which, in all likelihood, will not occur until sometime after January 1, 1979. However, this provision reflects a Congressional commitment to combat age discrimination. Such a commitment can be cited by the elderly and their advocates even before the legal provisions regarding age discrimination in revenue sharing programs actually take effect.

D. Public Hearings

The original Act made no provision for public hearings relative to the planned use of revenue sharing funds. The recent Amendments require governmental units to hold public hearings to provide citizens an opportunity to offer written and oral comment on the possible uses of revenue sharing funds. The governmental units are also required to publish, at least 10 days prior to the public hearing, in a newspaper of general circulation, a notice stating the proposed uses of revenue sharing funds and indicating the time and place of the public hearing. The Amendments express a strong intent that proper consideration be given to senior citizens and their organizations, since the following sub-section has been added to Section 121 of the original Act.

"(g) PARTICIPATION BY SENIOR CITIZENS -- In conducting any hearing required under this section, or under its own budget processes, a State or local government shall endeavor to provide senior citizens and their organizations with an opportunity to be heard prior to the final allocation of any funds provided under the Act pursuant to such hearing."

E. Review of Revenue Sharing Expenditures

The Amendments continue a requirement in the original legislation that jurisdictions which receive and expend revenue sharing funds issue a yearly report setting forth the amounts and purposes for which

funds received from revenue sharing have been appropriated, spent, or obligated during the entitlement periods, and showing the relationship of those funds to the relevant functional items in the government's budget. This report must identify differences between the actual use of funds received and the proposed use of such funds. This report precedes the public hearings scheduled on the proposed use of funds, and serves as a guide to the future use of funds.

4. Strategies

The provision allowing use of revenue sharing funds for matching, coupled with the added provision for participation by the elderly in the preliminary action involving allocation of funds, place older Americans in an advantageous position in their quest for revenue sharing funds. However, the elderly and their spokespersons must be actively and aggressively involved in the entire process whereby revenue sharing funds are allocated if proper advantage is to be taken of those opportunities. This makes it incumbent upon all in the aging network -- Regional Offices, State Offices, Area Agencies on Aging, agencies and organizations which receive Older Americans Act grants and contracts, and organizations of elderly persons -- to seek and find opportunities to assure adequate presentation of the needs of programs for the elderly to receive revenue sharing funds.

A. Hearings

Full advantage should be taken of the public hearings concerning the proposed use of revenue sharing funds. Network members may find the following suggestions helpful in exploiting the opportunities afforded by the public hearings:

- ° Aging agency personnel, advocacy groups representing the elderly, and other interested organizations should formulate a list of priority uses for revenue sharing resources well in advance of the local jurisdiction's formulation of its revenue sharing program.

- ° In formulating their recommendations, area agencies and organizations should give particular consideration to needs for non-recurring expenditures, such as development of senior centers, the purchase of furniture and equipment for aging programs, etc., since experience has shown that State and local governments prefer to use revenue sharing funds for such purposes rather than for programs which will require an on-going commitment. Opportunities for use of revenue sharing funds for matching purposes should also be explored in this process.
- ° The priority list should be shared with the decision makers through letters, briefings, and informal conversations. Every effort should be made to provide a detailed rationale, with as much supporting data as possible, for each request. This will be far more effective than simply presenting a list of demands.
- ° In contacting local officials about proposed revenue sharing expenditures, the elderly and their representatives should also offer suggestions about how best to schedule the hearing to facilitate participation by the elderly, as required under the 1976 Amendments. Special attention should be given to such key issues as use of facilities accessible to the elderly, a convenient time, etc.
- ° At the hearing itself, spokespersons can both commend the public officials for any favorable treatment the requests voiced by the elderly may have received and also renew requests which have not yet received favorable consideration.
- ° A written statement should be prepared in advance of the hearing to be submitted for the purpose of confirming the oral testimony.

- ° The use of a limited number of well prepared and well rehearsed spokespersons is preferred to a procession of individual witnesses.
- ° A good attendance by the elderly will show their interest and their support for their spokesperson. They should also encourage news media coverage of the hearings in order to gain as much publicity as possible.
- ° Following the hearing the spokesperson(s) should send letters to the participating public officials thanking them for their interest and attention and reiterating any requests which may yet receive a positive response.

B. Review of Previous Expenditures

This activity is necessary to effective advocacy of the elderly's interest in revenue sharing.

- ° Area agencies and other organizations should carefully assess each report published by the local jurisdictions concerning the use of revenue sharing funds. This can provide much information relevant to the formulation of new requests and proposals for projects of benefit to the elderly, to be presented during future funding cycles.
- ° Any failures to honor commitments to uses of funds for aging programs should be noted and called to the attention of the appropriate public officials. If possible an accounting for the discrepancy should be obtained and a commitment secured from the officials to take corrective action in the next revenue sharing expenditure cycle.

C. Nondiscrimination

Although the prohibition against discrimination on the basis of age will not become effective until January 1, 1979, or thereafter, full advantage

should be taken of the expression of intent in the 1976 Amendments against such discrimination.

- ° In formulating requests for specific revenue sharing allocations, area agencies, et al., should call the attention of the relevant officials to the fact that the anti-discrimination provisions will soon be in effect. These officials should be made aware that favorable consideration of requests by the elderly prior to the effective date of the anti-discrimination provisions would be an especially effective way of demonstrating good faith.
- ° When the effective date of the anti-discrimination provisions is established, area agencies and/or other organizations should take the initiative in formally notifying local officials of that specific date. This will serve as a reminder that the elderly require serious consideration in the allocation of revenue sharing funds.


5. Future Actions by AoA

On January 10, 1977 the U.S. Department of the Treasury issued final regulations pertaining to public involvement in decisions about the allocation of general revenue sharing funds. The requirements specified in the new regulations are reflected in the material transmitted via this Technical Assistance Memorandum. AoA will advise the network of any subsequent issuances that are specifically relevant to the process of securing a share of the general revenue sharing resources for the elderly. AoA will also advise the network when the Age Discrimination Act takes effect.

INQUIRIES: State Agencies should address inquiries to Director, Office of Aging, Regional HEW Office.

Area Agencies on Aging should address inquiries to State Agencies on Aging.

Title VII Nutrition Projects should address inquiries to their grantor (State or Area Agency on Aging).



 Arthur S. Flemming
 Commissioner on Aging

DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE
WASHINGTON, D.C. 20201

OFFICIAL BUSINESS
PENALTY FOR PRIVATE USE, \$300

POSTAGE AND FEES PAID
U.S. DEPARTMENT OF H.E.W.

391



FIRST CLASS

Mr. Daniel Schaulder OSCP-3
Special Assistant to the
Governor on Aging
504 State Street Bldg.
Harrisburg, Penn. 17101

CENSUS TRACT PAPERS

Series GE-40

- No. 1. Papers Presented at the Census Tract Conference, September 10, 1965, Philadelphia, Pa.
- No. 2. Needs and Plans for the 1970 Censuses (Census Tract Conference of August 15, 1966, at Los Angeles, Calif.)
- No. 3. Some Uses of Census Tracts in Private Business, by Wilbur McCann (1967)
- No. 4. Papers Presented at the Conference on Small-Area Statistics, American Statistical Association, Washington, D.C., December 27, 1967, and Related Papers
- No. 5. Papers Presented at the Conference on Small-Area Statistics, American Statistical Association, Pittsburgh, Pa., August 23, 1968, and Related Papers
- No. 6. Final 1970 Census Plans and Four Programing Systems for Computerized Data Retrieval and Manipulation (Conference on Small-Area Statistics, American Statistical Association, August 21, 1969, New York, N.Y.)
- No. 7. New Uses of Census Resources, The Southern California Census Use Study and a Related Paper (Conference on Small-Area Statistics, December 29, 1970, Detroit, Mich.) American Statistical Association.
- No. 8. Small-Area Statistics: Strengthening Their Role in Federal Government and Their Use in Criminal Justice Programs (Papers Presented at the Conference on Small-Area Statistics, American Statistical Association, August 23, 1971, Ft. Collins, Colo.)
- No. 9. Social Indicators for Small Areas, Papers Presented at the Conference on Small-Area Statistics, American Statistical Association, August 14, 1972, Montreal, Canada.
- No. 10. Statistical Methodology of Revenue Sharing and Related Estimate Studies (Papers Presented at the Conference on Small-Area Statistics, American Statistical Association, December 27, 1973, New York, N.Y.)

Copies of early editions may no longer be in print but are usually available in Federal Depository Libraries. Papers presented at conferences held from 1958 through 1964 were originally published by the Bureau of the Census in its Working Paper Series. For information on all census tract reports, write to:

Subscriber Services Section (Publications)
Social and Economic Statistics Administration
Washington, D.C. 20233

U.S. GOVERNMENT PRINTING OFFICE
DIVISION OF PUBLIC DOCUMENTS
Washington, D.C. 20402
OFFICIAL BUSINESS
FIRST CLASS MAIL

PENN STATE UNIVERSITY LIBRARIES



A000072860175

PRINTING OFFICE
375

U.S. MAIL

